Context Effects and the Structure of Concepts

Karen Lyon

Ph.D.

University of Edinburgh

1989

I declare that this thesis was composed by myself, and that the work reported here is entirely my own, unless otherwise indicated.

ACKNOWLEDGEMENTS

This research was funded by an SERC award (no. 85310119) to Karen Lyon.

I gratefully acknowledge the help of James Hampton in providing his original stimuli and for his helpful comments and suggestions. Thanks are also due to Douglas Medin and several anonymous reviewers who commented on experiments submitted for publication. I am indebted to Frances Provan for her statistical advice, and to Malcolm Lyon for formalizing the proof. I benefitted much from discussion with folk from Cognitive Science, and would particularly like to thank members of the Mental Lexicon Workshop. Thanks are also due to my supervisor, Terry Myers, who gave me tremendous freedom in my research. Finally, I would like to thank Simon Garrod and Brendon McGonigle who examined this thesis for their advice and suggestions.

Most of all I would like to thank all my friends who encouraged me, put up with me, scolded me, distracted me, and who eventually saw me through this experience. Thanks to Nick Chater for his boundless optimism, startling insight and terrible stapling, with whom it was a privilege to work.

Lastly, thank-you very much to 236 subjects from Edinburgh University, Napier Polytechnic, St. Paul's & St. George's Episcopal Church and Edinburgh Youth Hostel, without whom this thesis truly would not have been possible.

"When *I* use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean – neither more nor less."

"The question is," said Alice, "whether you *can* make words mean so many different things."

"The question is," said Humpty Dumpty, "which is to be master – that's all."

Lewis Carroll, *Through the Looking Glass*

Context Effects and the Structure of Concepts

Contents

5. WHY OVEREXTEND: COMPENSATION BETWEEN DIMENSIONS?

6. CATEGORIES AND CONCEPTS: A FUTURE PERSPECTIVE

NOTES

REFERENCES

APPENDIX 1

APPENDIX 2

APPENDIX 3

APPENDIX 4

Abstract

This thesis investigates the effects of contextual change on conceptual representation, and in particular the changes involved when concepts combine. In Chapter 1 an overview of the Classical Approach, the Probabilistic Approaches and the Exemplar Approach to category representation is given. Each approach is examined with regard to four conceptual functions: stability, representation of ontological and linguistic meaning, simple categorization and complex categorization. It is concluded that these theories vary in the extent to which they can fulfil the first three functions, but that none can adequately address complex categorization. This highlighted the need to understand concept combination more fully.

In the second chapter a normative study of conceptual stability is reported. It was found that changes in context cause a complete change in the goodness of example of instances. In Chapter 3 two experiments are reported which investigate adjectival qualification of noun meaning while reading. Results suggest that context immediately affects the retrieval of nouns, casting doubt on the role of category prototypes.

In Chapter 4 four experiments are reported which investigate noun-noun combination. In particular, the question of whether nouns combine according to a Boolean model is assessed. It is concluded that subjects do not use Boolean combination, but tend to overextend conjunctive concepts. Implications of this finding are discussed with reference to Hampton's Inheritance of Attributes model.

In the fifth chapter an alternative explanation for overextension is introduced and tested empirically. The results of this experiment clearly demonstrate that the more categories subjects have to combine the greater their tendency to overextend. This finding, consonant with the proposed hypothesis that subjects use a best fit strategy for deciding category membership, is compared to the

Inheritance of Attributes model.

In the final chapter traditional theories of conceptual representation are reassessed. It is proposed that the localist nature of these theories, where concepts are conceived of as isolable units, is inappropriate for adequately capturing the knowledge-centred nature of concepts. A globalist position where concepts are generated on-line is outlined, and possibilities for distinguishing the two approaches are discussed. Globalism is seen as an appropriate theoretical framework within which to develop theory based approaches to conceptual representation.

CHAPTER 1

A THEORETICAL RESUME OF CONCEPTS AND CATEGORIZATION

1.1. Towards a Definition of Concepts

A concept in very general terms might be construed as a stored representation of an object, action, or notion, which can be grasped or considered by the mind and is manifested in the language system. This description, however, is not sufficiently precise as a basis for experimental investigation. How can it be decided if someone has a given concept or not; which concept a person is using in a given situation; and which level of language (e.g. words, typologies) reflects conceptual representation? A more precise definition of this important psychological notion is unavailable for good reason, as Kendler aptly put it:

> After all, a concept is not a single thing or event whose attributes and functions can be simply listed and described. A concept is a complicated psychological phenomenon, a complete description of which will be contained someday in a theory capable of explaining the numerous empirical laws involving the term concept.

Kendler, 1964) p 226.

Given the understanding that the notion of a concept must be a complex one, and indeed one which cannot yet be fully defined, it is nevertheless necessary to adopt a pertinent working definition if the acquisition of meaning is to be studied at all. In practice, concepts are taken to be "the general idea or meaning which is associated with a word or symbol in a person's mind... the abstract meanings which words and other linguistic terms represent" (Longman Dictionary of Applied Linguistics). Different researchers and different disciplines expect concepts to fill different roles. Rey (1983) has highlighted four roles concepts have been invoked to perform. Not everyone agrees that in an adequate theory of cognition, concepts will fill all four roles. What is clear, is that every conceptual theory will have to take each role into account. In order to clarify the explanatory scope, and possible adequacy, of several conceptual theories, they will be assessed with regard to Rey's four conceptual functions.

Rey's Four Conceptual Functions

1. Stability Functions : Concepts enable the formation of stable cognitive states in response to particular stimuli. This is a necessary prerequisite for communicating with oneself and others.

2. Representation of Ontological / Linguistic Meaning : Concepts represent the whole meaning of a term including some relational information (for example, semantic equivalence and antonymy relations). A concept also provides information on which logical inference is based. The concept BACHELOR (Note 1) *necessitates* being unmarried because of the information contained in the concept. It is also this information which allows us to draw semantic implications. For example, if we know the statement "Jane likes apples" is false, then that implies the statement "Jane likes all fruit" must also be false, because of what we know about apples and fruit.

3. Simple Metaphysical and Epistemological Function (Note 2): Concepts are the representation of the information which allows us to make simple categorization judgements. Concepts also contain information about how such judgements are carried out.

4. Complex Metaphysical and Epistemological Function : Concepts (or complex concepts) are representations of the information which allows complex categorization judgements (e.g. is "haddock" a PET FISH?), and the means by which such categorization is carried out (e.g. comparison with the prototype of a complex concept PET FISH).

It is apparent that the primary role of concepts is one of representing the world to oneself and others. The other functions concepts serve are subsidiary means to this final end, but subsidiary means have proved easier to investigate. Thus, an overwhelming proportion of researchers have chosen to study the third conceptual function listed, that of simple categorization processes. I will discuss the findings and implications of this research in some detail. This chapter is devoted to assessing which of Rey's conceptual functions current theories of concepts address. In section 1:3 I will consider whether complex categorization can be explained by models which have developed from phenomena associated with simple categorization. Problems associated with the second conceptual function listed – the representation of whole meaning – will arise throughout the course of the discussion but will be highlighted in section 1:4. The posited stability of concepts will be seen to vary somewhat with the particular view of concepts one adopts; the question of whether the onus of providing stability should be placed on concepts will be raised later in the thesis (see Section 3:1).

Three cognitive theories: the Classical View, the Probabilistic Views and the Exemplar View, represent the basic ideas underlying individual conceptual theories. I will assess the strengths and weaknesses of each approach with regard to the four roles concepts might play. Simple categorization has been studied in the hope that some knowledge about what criteria are brought to bear in categorization decisions, will lead to a greater understanding of conceptual representation. Categorization will be examined through the eyes of each cognitive theory, and in turn these cognitive theories will face the test of categorization phenomena. In particular, the inability of current theories to address themselves adequately to complex categorization decisions will be highlighted. This chapter concludes with a discussion of conceptual combination.

1.2. The Classical View of Concepts

Assumptions

What has come to be called the Classical View of concepts has been handed down to psychologists from philosophy, and is notably Aristotelian in origin. This view is primarily concerned with what has been called the intension, core meaning, or metaphysical aspect of concepts. Firstly, the Classical View assumes a unitary, summary representation of an entire class, where the summary assumption is justified on the grounds of limited processing capacity. The second, and distinguishing assumption of the Classical View states that the features which make up the summary representation are defining for the concept; that is to say, they are singly necessary and jointly sufficient to define the concept. Thus, in any given categorization task the summary representation will be retrieved, then compared with the item in question. If the summary representation maps on to the item it is deemed a member of the category, otherwise it is rejected. Adopting the notion of defining features, leads to the prediction that the summary representation of a category must, by definition, contain fewer features than any given category instance because it is the common denominator of the category. The third basic assumption of the Classical View is logically implied by the first two: namely, a concept which is a subset of another concept will have the defining features of the superordinate concept nested within it.

e.g.	CONCEPT	HYPOTHETICAL DEFINING FEATURES
	Quadrilateral	A polygon, Four sides
	Parallelogram	A polygon, Four sides, Opposite sides parallel, Opposite sides equal
	Rectangle	A polygon, Four sides, Opposite sides parallel, Opposite sides equal, Four right angles
	Square	A polygon, Four sides, Opposite sides parallel, Opposite sides equal, Four right angles, Four equal sides

The implication that superordinate categories should have fewer defining features than subordinate categories has some experimental parallels e.g. Rosch, Simon & Miller (1976) found that subjects could list fewer features of superordinates than basic-level instances.

A CLASSICAL PERSPECTIVE ON REY'S CONCEPTUAL FUNCTIONS

Endorsing the basic assumptions of the Classical View leads to specific predictions about conceptual attainment and usage in a wide variety of semantic tasks. In order to convey an impression of the applicability of the theory as a whole, I will summarize its basic assumptions, and the arguments levelled against them, with respect to Rey's four conceptual functions. I will then consider more severe challenges to the basic tenets of the theory.

1. Stability

The Classical View implies the complete stability of concepts once they are formed. That is to say, when presented with a breed of dog one has never encountered before it will never be necessary to change one's representation of dog since the novel dog must, by definition, already match the summary representation of the class. The Classical View thus provides a basis for interpersonal and intrapersonal stability since each concept has a definition, and is thus the same on all occasions of use. In this way Classicalists can also expect concepts to be productive. That is, a few stable, basic units should be able to join together predictably to produce new stable concepts, in much the same way as the same words can join together to form sentences with different meanings.

Some developmental flexibility is, of course, allowed, and the Classical approach to concept acquisition is worth briefly outlining here. Eve Clark (1973a) has taken such an approach, namely, that children acquire stable concepts through a process of "component by component" acquisition. That is to say, the defining featural properties by which adults represent words (according to the Classical View) are acquired one-by-one by the child. Thus, a child's lexical entry for a word will be *an incomplete version of an adult's lexical entry* for the same word. Eve Clark supported this claim with developmental evidence showing the apparent synonymy of antonyms like "more" and "less" at a stage in concept attainment. Moreover, the meaning assigned to both these words was the meaning of, what was claimed to be, the less featurally complex, unmarked item. This work has since been criticized methodologically and experimentally by Susan Carey (1982). Carey pointed out that these effects were found to reduce if the data for individual children was analysed separately. She also suggested that error rates may be an immature reflection of adults' increased reaction times, hence children may have been making the same distinctions between positive and negative terms as adults. Lastly, Eve Clark (1973b) also found that the concepts "in", "on" and "under" were used synonymously. In this case, however, the selection of which word was learned first, and which meaning was dominant, could not be explained in terms of featural simplicity. Thus, it would seem that children do not simply learn adult concepts feature by feature. Susan Carey has suggested that children form their own concepts and theories about what words mean, rather than partial adult concepts. If this is the case, then a Classical explanation of interpersonal and intrapersonal stability is difficult to achieve. It seems that the Classical approach no longer has the explanatory edge when describing the developmental growth of stability.

2. Representation of Ontological/Linguistic Meaning

The second function which concepts are invoked to perform is the representation of the whole meaning of a term. *Prima facie*, the Classical View can cope easily with phenomena that cause problems for other theories. For example, the fact that being unmarried is entailed by being a bachelor is because <-married> is one of the features of BACHELOR. Similarly, semantic implication can be explained with recourse to the semantic features of a word. We know that if the statement "Tom's wife has died" is true, it implies that the statement "Tom never married" is false because the feature <+married> is present in the antecedent. Is this, in fact, a satisfactory *explanation* of semantic implication?

Granted, the Classical View can explain *some* semantic implication at the word level in terms of features: for instance, it might be inferred that sickness or death would result from drinking poison since poison contains the feature <-potable>. In general, however, this strategy merely displaces the problem to a featural level: the question "Why can't men drink poison?" is replaced with questions like "Why does <+male> imply <+animate>, and the combination of <-potable> and <+animate> spell disaster"? What recourse has a Classicalist in this case? Perhaps it would simply be argued that today's features are merely postulates, and not the indivisible primitives which really constitute conceptual building blocks. I would like to offer two comments on this point. Firstly, the existence of semantic primitives was originally proposed in response to the limited processing capacity of the brain (Johnson-Laird, 1983) it now seems they are creating the burden of a huge analysis for each word. Secondly, it is difficult to conceive of a more primitive level for features such as <+male> and <+animate> that could have any discriminatory value. It seems a more parsimonious suggestion that we use words (rather than features or primitives) to communicate precisely because they are pitched at an optimum level for our cognitive processes (a point conceded by Miller (1978) by asserting that normal processing uses a gestalt).

Explanations resorting to features can often seem inadequate: for example, explaining antonymy in terms of featural analysis may lead one to posit features in each antonym which are positive and negative poles of a dimension. This is, on reflection, no more an explanation of the relations between words as it would be to say that words are antonyms because they are dimensionally opposed! In conclusion, featural explanations of connections between words are ad hoc, they merely displace the problem to another level rather than addressing it directly.

3. Simple Categorization

The Classical View states that categorization is achieved by comparing the item to be categorized with the summary representation of the class. Therefore, it should be possible to categorize all the instances of the concept once it has been fully obtained (hence there should be no ambiguous instances like "tomato" (fruit), "rhubarb" (vegetable) and "lightshade" (furniture)). It should also be as easy to categorize a common instance as a novel instance, the process being the same in each case. These predictions have been repeatedly contradicted by people demonstrating the effects of 'goodness of example' on categorization tasks (e.g. Rosch, 1975a, 1978; - see Medin & Smith, 1984 for a review). This criticism is partly redressed by Rey who claims that concepts have a metaphysical role, therefore data relating to epistemological classification is irrelevant to the internal structure of the concept - see section 1:5.

4. Complex Categorization

Accepting a definitional view of concepts lends itself to Boolean descriptions of conceptual combination like those found in natural set theory, where membership is all-or-none. Such a theory predicts that similarity relations between concepts will be symmetrical. Rosch (1975b) showed, however, that good rather than poorer category members were used as 'cognitive reference points'. Hence, people find it natural to say: "A zebra is like a horse" not "A horse is like a zebra". Zadeh (1965) has emphasized that natural class concepts are not accurately described by the all-or-none membership characterized in natural set theory.

The inability of the Classical View to provide an account of complex concepts points uncomfortably to its limited scope as a conceptual theory. It may be that many of the Classical View's apparent problems with complex categorization could be avoided by taking Rey's insistence on the metaphysical/epistemological distinction seriously. It implies that we should expect Boolean combination to be intensional not extensional. That is, membership in a complex concept should be a Boolean combination of the conceptual criteria for membership in the constituent categories, not the intersection of the sets of members of those categories. The necessity of adopting an intensional approach to combination has been argued by James Hampton (Hampton, 1987a) - see also section 4.6. One problem with such a view is that the theoretical notion of intensional combination is untestable while there is no way of discerning the features present in a concept.

Criticism of the Basic Assumptions of the Classical View

The three basic assumptions of the Classical View as outlined above will be discussed with respect to data obtained from simple categorization studies. Firstly, the assumption "a concept which is a subset of another concept will have the defining features of the first nested within it", implies that ROBIN and CHICKEN should be equally easy to verify as instances of BIRD, both having the defining features of BIRD nested within them, and being at the same level of abstraction. Yet it has been consistently found that CHICKEN and other non-typical examples take longer to verify. Similarly, the statement "A canary is a bird" can be verified more quickly than "An ostrich is a bird". In order to get round this problem, Rips, Shoben & Smith (1973) and Smith, Shoben & Rips (1974) have proposed that concepts contain *defining* features and *characteristic* features. Initially, the total number of features for CANARY would be accessed and globally compared with the features of BIRD enabling a rapid response if the degree of featural overlap was great. In the case of judging "ostrich" as a member of "bird", the total number of shared features would be less (since OSTRICH does not share many characteristic features with BIRD), hence necessitating a second stage defining feature analysis. This view, however, is tantamount to rejecting that "all instances possessing the criterial attributes [of a category] have a full and equal degree of membership" which is the basis of the Classical View (Rosch, 1975a). The remaining question is why it is necessary to propose a complicated two stage analysis, when comparing only the defining features would give completely accurate results?

A second way to solve the problem of some categorization decisions being easier than others, is to endorse the ad hoc assumption that good examples of categories have fewer features than poor examples. Thus, a good example will be more quickly verified as a category member because fewer features have to be searched before finding the defining features mapping the summary representation.

This approach also runs into problems, however, when forced to explain the priming effects found by Rosch (1975a). Rosch asked subjects to decide whether two instances were members of the same category. In one condition the subjects were warned of the onset of the pair, by a signal. In a second condition, subjects were primed with the category name. Rosch found that priming facilitated decision speeds for good examples like "apple-apple", and *hindered* decision speed for poor examples like "marrow-marrow" with respect to the signal control. The postulated larger number of features in the poor examples does not provide any explanation of why the decision process should actually be slowed down when a prime, rather than a signal, was used. What is quite clear is that without major theoretical changes the Classical View cannot account for the prototypical structure that so many have reported (Wittgenstein, 1953; Rosch, 1973, 1975a, 1978; Rosch & Mervis, 1975; Glass & Holyoak, 1975; Garrod & Sanford, 1977; Tversky & Gati, 1978; McCloskey & Glucksberg, 1979; Roth & Shoben, 1983).

Simple categorization also provides a more fundamental challenge to the Classical View: a challenge to the notion of any defining features. If there are defining features, given enough time subjects should be able to correctly classify items; yet many ambiguous items remain. A Classicalist may well respond that these people could not have fully attained the given concepts, if this is true it merely points to the idealized nature of the Classical theory of categorization! Finally, as Wittgenstein (1953) illustrated in his games example, it is not know, nor is it apparent how it could be discovered, what the features, and defining features in particular, of concepts are.

The Classical View is thus forced to amend two of its basic assumptions, and propose that the summary representation of the class may contain the defining features of the class *plus* some typical characteristic features of the category. This altered position bears a remarkable family resemblance to the probabilistic approaches outlined in the next section. The remaining assumption of the Classical View is that categories are represented by a unitary summary representation. This assumption has not been directly challenged by work on simple categorization, and is also adopted by most of the Probabilistic Views of conceptual representation. It has not been left totally unchallenged, however, as shall be seen in section 1:4.

1.3. The Probabilistic Views

The alternative theories which grew out of the categorization findings have been classified as the Probabilistic Views. They have in common the replacement of the problematic notion of defining features, with a probabilistic membership criterion. As a result, the summary representation of the class is conceived of as being more complex than it appears to be in the Classical View. The summary representation of a class incorporates a number of salient characteristic properties previously excluded because of a few outliers e.g. penguins cannot fly hence <+fly> cannot be a necessary feature of BIRD. According to the Probabilistic Views, categorization of an instance is accomplished by exceeding a critical probabilistic similarity weighting. There are two major variations of the Probabilistic perspective: the Featural approach and the Dimensional approach.

<u>The Featural Approach</u>

The assumptions of this view as characterized by Smith & Medin (1981) are: (i) concepts are represented by a summary description of the class, and (ii) the features which make up that summary description are salient ones, with a high probability of occurring in instances of the concept. Note that this summary description may not equal, or even map on to, any *particular* instance of the concept, but it will be more similar to some instances than to others. It is this latter factor that is taken to account for typicality findings. The general assumption about the criterion for categorization is:

$$\exists x: x \in Y <=> F(x) > C(Y)$$

Where x and Y are concepts, F(x) is a featural weighting function on x, and C(Y) is a critical weighting for the concept Y.

Many researchers have adopted a general featural approach to categorization (Collins & Loftus, 1975; Hampton, 1979; McCloskey & Glucksberg, 1979; Hayes-Roth & Hayes-Roth, 1977). The approach gives an increased role to discriminatory, but non-universal category features; this change legitimizes the over-generalization so typical of human thinking e.g. "fruit is sweet". There are also problems with the approach, the Probabilistic Featural account, like the Classical View, is dogged by the immense problem of reliance on unknown features. When the total explanatory power of a theory is based on the distribution of features within a class, it is obvious that that theory is weak while we know neither that distribution nor the features from which it is composed.

The Dimensional Approach

The Dimensional approach discards the traditional notion of featural components entirely. Instead it proposes that concepts are characterized by values along a number of *dimensions*. For example, APPLE is said to be closer to the concept FRUIT than GRAPE because its salient dimensional characteristics (like <size>) are closer to the category mean. The fact that the Dimensional Approach relies on quantitative rather than qualitative features (Garner, 1978) may not seem an important difference, but the quantitative approach has led to a new way of describing concepts. It has been argued that concepts containing the same relevant dimensions can be represented in a multidimensional metric space (Shepard, 1962, 1974; Henley, 1969; Carroll & Wish, 1974; Homa & Silver, 1976). A metric space is defined as a set M together with a function $\delta: M \times M \rightarrow R$ where δ exhibits minimality, symmetry and triangular inequality i.e. for any a and b members of M:

- (i) $\delta(a,b) \geq \delta(a,a) = 0$
- (ii) $\delta(a,b) = \delta(b,a)$
- (iii) $\delta(a,b) + \delta(b,c) \geq \delta(a,c)$.

In this approach "semantic distance" across the metric space is analogous to degree of similarity between words, and is therefore a useful way of representing the categorization process:

$$\exists x: x \in Y \Longleftrightarrow D_{(p-x)} < S(Y)$$

Where x and Y are concepts, D represents semantic distance, p is the central tendency of the concept, and S(Y) is a critical semantic distance for the concept Y).

Thus, in the Dimensional approach, there was an adoption of Boolean models, the assumptions underlying which proved such a problem for the Classical View. The assumption that conceptual space must be metric has no empirical base; a growing number of researchers found experimental justification for its rebuff (Beals, Krantz & Tversky, 1968; Tversky & Krantz, 1970; Krantz & Tversky, 1975; Boyd, 1972). For instance Tversky & Gati (1978) showed that similarity judgements are not associative (contradicting point (ii) above), and Arnold (1971) showed that dissimilarity judgements are not Euclidean (contradicting point (iii) above). These problems have to some extent been recognized and multidimensional scaling techniques are increasingly being used in non-metric spaces.

Dimensions Versus Features

Although the Dimensional approach may be an adequate way to represent colour, size, and other intuitively quantitative concepts, it does not seem appropriate for all concepts. As Smith & Medin (1981) have pointed out "many properties lack dimensional aspects like continuity and betweenness". It seems certain that some concepts, for example those of faces, are better represented featurally. More evidence for a Featural account is provided by Goldman & Homa (1977) who argue that normal summary representation is most likely to be modal in nature, whereas it seems most likely that dimensions are represented by averages. This is because it is unlikely that a dimensional quantity, such as size, will cluster on a specific value such as 5'9". Rosch & Mervis (1975) pointed out that subject reports of word constituents support a Featural over a Dimensional representation. Finally, Smith & Medin (1981) have pointed out that dimensions can always be represented featurally (quantitative changes can be perceived as qualitative changes) but not vice versa. Hence, if given a list of women's heights, heights up to 5'3" could be represented by the feature <short>, from 5'3" to 5'7" by <medium height>, and over 5'7" by <tall>. If given these three features, however, they could not be recoded dimensionally.

A PROBABILISTIC PERSPECTIVE ON REY'S CONCEPTUAL FUNCTIONS

Having thus outlined the nature of the Featural and Dimensional approaches to conceptual representation, it is important to assess the merits of the Probabilistic View as a whole. In particular, I will discuss the View's explanatory capability with respect to Rey's conceptual functions.

1. Stability

The Probabilistic View depicts concepts as flexible. Thus, a person will never reach a stage of "conceptual attainment" where classification errors will be zero. The nature of the approach characterizes concepts which change and develop as experience changes and develops. From a developmental perspective, it can be predicted that children's concepts should be more variable than those of adults because children encounter more novel, incompatible instances to elicit such changes. It is also implied that children do not have "incomplete versions of adult concepts" but that they form their own concepts according to experience. Although it cannot be said that adults have complete conceptual stability, in the sense that their concepts are lexically frozen, it is certainly thought that adults reach conceptual maturity or competence.

An important distinction between the Probabilistic and the Classical View with respect to stability, is that the Probabilistic approach looks for stability *outside* conceptual definitions. Firstly, stability results from experience of conceptual usage. Thus, by simple virtue of the fact that the more often you encounter a concept being used, and interpret its meaning successfully, the less likely you will be to encounter a novel use which causes you to alter your concept. Secondly, stability results from the perception of a structured world, or innate cognitions which impose structure on our environment. This assumption is not as dramatic a difference from the Classical View as it might at first seem. The Classicalist avoids this philosophical debate only by referring to an intermediate cause of stability viz defining features. If pressed for a more complete explanation of stability many Classicalists would propose that the terms of definitions, or primitive features, are innate (Fodor, 1981).

2. Linguistic

The Dimensional and Featural accounts deal differently with the linguistic functions of concepts. The Dimensional account explains relations of synonymy in terms of minimal separation in semantic space. Although sufficient for synonymy this type of explanation does not suffice for antonymy because it cannot account for the relationship that distinguishes antonyms from other dissimilar words. That is to say, postulating that antonyms are maximally dissimilar does not account for the psychological connection between opposites. Semantic implication is also inadequately captured by referring to semantic distance alone: it does not seem feasible to suggest that all words that can be involved in a process of semantic implication are located together in semantic space e.g. the feature <+animate> is involved in hundreds of different semantic implications between word pairs. Therefore it seems necessary to posit a relationship between words that is not solely dependent on similarity: that, by definition, is outwith the scope of this approach. The Featural account explains linguistic function in a similar way to the Classical View: that is, synonymy and semantic implication can be explained in terms of underlying conceptual features. The difference with the Featural account is that the underlying features are not defining, but characteristic. Antonymy is also accounted for in terms of features, that is a concept X exceeds some critical value of dissimilarity for a concept Y. Yet again the explanation of antonymy is couched in terms of maximal dissimilarity - an explanation which fails to capture the salience of this relationship.

3. Simple Categorization

As the Probabilistic View was formulated in response to simple categorization data, it explains typicality very well. For the Featural account, the closer an instance is to the modal summary representation of the class, the easier it is to categorize. For the Dimensional account, the closer an instance is in semantic space to the average summary representation of the class the easier it is to classify. Similarly, because membership is not "all or none" some instances have unreliable classifications. This explanation of poor members relies on criterial membership judgements, is this tantamount to specifying a category boundary by the back door? I would argue that it is not: a criterion for category membership is comparable to a judgement confidence level which will vary from individual to individual, and in the same individual, over time. The criterion level is subject to the same kind of influences as described in Signal Detection Theory (Egan & Clarke (1964)). For example, rewards for positive classification will cause subjects to be more lenient when applying membership criteria, whereas punishment for false inclusions will cause them to become stricter.

4. Complex Categorization

Until around 10 years ago, work on concepts and categorization had largely excluded complex categorization from serious study. The remark: "what holds for simple categorization... need not hold for other functions" made by Smith & Medin (1981) has proved to be prophetic. The probabilistic models can provide no clear idea as to how the concept PET FISH is derived (if it is) from the concepts PET and FISH. To examine this problem more thoroughly I will look at the most developed theory Probabilists adopted as a way of answering these questions: that of Fuzzy Set Theory.

Fuzzy Set Theory was originally proposed by Zadeh (1965). It provided a theory capable of representing relations between sets that were not all-or-none. The Fuzzy Set approach was adopted by a number of researchers (Lakoff, 1973; Oden, 1977; McCloskey & Glucksberg, 1978) but it has since been criticized as an inadequate expression of complex concepts (Osherson & Smith, 1981; Roth & Mervis, 1983). Osherson's argument is aimed at highlighting the inadequacy of the "prototype" view (a general term for either of the Probabilistic approaches or the Exemplar View).

Osherson defines complex concepts as the resultant concept when "one or more concepts combine together to form another whenever the latter has the former as constituents". Usually this is reflected grammatically in the forming of a word pair e.g. "pet" and "fish" combine to form the complex concept PET FISH. What is of interest is the relationship between a complex concept and its simpler constituents. Osherson describes Fuzzy Set Theory's representation of conceptual combination as the intersection F such that:

$$\forall x:(x \in F)(C_{\text{striped apple}}(x) = \min(C_{\text{striped}}(x), C_{\text{apple}}(x))$$

That is, if an instance "striped apple" was rated a good example of APPLE and a bad example of STRIPED, it would be rated a bad example of STRIPED APPLE, as an instance is never more representative of a complex concept than it is of any of its constituents. Fuzzy Set theory, then, does not capture the intuitive interaction between constituents when they form complex concepts. It seems to me that the term "complex concepts" may subsume two kinds of word combination, each of which could be expected to bear a different kind of relationship to its constituents. That is, when two words join together *interactively* to form a new concept, Fuzzy Set Theory cannot describe the relations between the constituents. Hence:

$$C_{\text{red fox}}(x) \neq \min(C_{\text{red}}(x), C_{\text{fox}}(x))$$

Membership in the category "Red fox" does not equal minimum membership in the constituent categories because "fox" alters our understanding of "red". Similarly:

$$C_{\text{dark horse}}(x) \neq \min(C_{\text{dark}}(x), C_{\text{horse}}(x))$$

Membership in the category "dark horse" is not the minimum membership of the constituents as "dark horse" in this context has a specific new meaning. One could think of many such examples. However, if the word pair "red car" is considered then the interaction intuitively no longer exists, thus:

$$C_{red\ car}(x) = \min(C_{red}(x), C_{car}(x))$$

Even in this case where there is no intuitive interaction between RED and CAR, there is no reason to suppose the minimality rule will hold out. Subjects might be likely to say that a "red car" is a better member of the conjunct than of either constituent simply because it meets the membership criteria for *both* concepts. Thus, it may be the case, that some concepts are compositional while other concepts are interactive. This distinction is certainly intuitively appealing, and one that merits some investigation.

Roth & Mervis (1983) provide an excellent empirical refutation of the appropriateness of Fuzzy Set Theory. Two predictions other than the intersection prediction, dealt with by Osherson, are those of containment and union. Both of these predictions are challenged by Roth & Mervis. Union is defined as the function on a member x such that:

$$fC(x) = \mathrm{Max}[fA(x), fB(x)]$$

Where C is the fuzzy set resulting from the union of the sets A and B, and f(x) is a membership function on each set.

Hence the membership of an instance in category C (an A or a B) is equal to its best membership rating in a constituent category. This formulation seems intuitively correct, and from it we can see how the fuzzy set notion of containment arises. The set called C which is the union of A and B can be used to demonstrate the special case of containment, that is the union of the sets A and B, when A is a subset of B. In normal set theory it is obvious that in this case C is equal to B. Similarly, this is the case in fuzzy set theory and so containment is defined as the function on a member x such that:

$$\forall x\ fA(x) < fB(x) <=> A \subset B$$

Where A and B are sets and f(x) is a membership function on each set.

So, whenever A is a subset of B, instances of A will have *at least* as high a membership value in B as in A. Thus we can predict "robin" should be at least as good an example of ANIMAL as it is of BIRD. Roth & Mervis found that this kind of prediction does not always hold, which implies that either our characterization of containment is inappropriate, or the set A in question cannot be a subset of B. Roth & Mervis collected goodness of example ratings of superordinate (animal) and subordinate (e.g. mammal, bird, fish) categories. Each exemplar was rated once as a member of its close superordinate category (e.g. fish) and once as a member of its distant superordinate category animal. Containment predicts that an exemplar should be at least as good an example of its distant superordinate as its close one; union predicts that an exemplar's representativeness in the distant superordinate will *equal* its representativeness in the close superordinate of which it is most typical.

Roth & Mervis found that these predictions did not generally hold. The level of representativeness varied among superordinates, but it was often the case that instances were more representative of close superordinates than distant ones; but in some cases the opposite was true. This means that the predictions of Fuzzy Set Theory are not upheld, unless it is true that close superordinates are not in fact subsets of distant ones. This demonstrates that the relationship between different levels in the conceptual hierarchy is not a simple one. Their study led Roth & Mervis to conclude that: "the Goodness of Example of an exemplar in categories at different levels within a taxonomy is determined independently". Such a conclusion suggests that the Fuzzy Set model of conceptual combination is not psychologically valid.

In conclusion, it can be seen that Probabilistic theories provide a more satisfying account of the data relating to simple categorization processes than the Classical View. However, the extension of a Fuzzy Set approach to complex categorization has been unsuccessful, and complex categorization remains a major problem for Probabilistic theories of concepts.

1.4. The Exemplar View

The Exemplar View is a major alternative to the Probabilistic Views, as it also provides an account of categorization which incorporates typicality effects. The factor which distinguishes the Exemplar View from the Classical and Probabilistic Views is its holistic nature. The Exemplar approach does not require the postulation of underlying features or dimensions in order to explain typicality effects. It is this view, in my opinion, that is closest to the notion of a prototype as first described by Elenor Rosch (Rosch, 1975a). Crucial to this view is the hypothesis that categorization is achieved by comparing the instance in question with the prototype, or best example, of that class (see Medin & Schaffer, 1978). The greater the similarity between the prototype and the instance the more easily that instance can be classified as a member of the category (Note 3). It is not necessary that the prototype be a single category member, it may well consist of a small group of best instances e.g. apple, orange and pear may be the prototypes for the category "fruit". This view has a number of theoretical advantages over alternative approaches. Firstly, it does not require the postulation of a category concept that is somehow abstracted from the members of that category: the category is mentally represented in terms of its most typical referents (Note 4). Secondly, the Exemplar View provides a simple explanation of why subjects remember individual instances, rather than the summary representation of the class. This is reported to be so even in tasks where the subject is only required to learn the criteria for categorization into a small number of groups (Smith & Medin, 1981). In a recent paper, Brooks (1987) has reported findings supporting the relevance of storing exemplars. He argues:

> our conceptual resources are considerably more *decentralized* than is usually portrayed, consisting in routine analogy to prior interpreted episodes as well as to more abstract principles and models.

Brooks, 1987 p. 142

Thirdly, the holistic approach of the Exemplar View provides good theoretical convergence on the salience of the basic level (Rosch, 1978).

A HOLISTIC PERSPECTIVE ON REY'S CONCEPTUAL FUNCTIONS

The Exemplar View provides a challenge to more traditional views of concepts. During the process of assessing the validity of such a challenge, I will look in turn at each of Rey's conceptual functions.

1. Stability

The Exemplar approach predicts a high level of conceptual stability because typical (good) exemplars are the most frequently encountered, and usually the first to be acquired. The approach also allows for the fact that the extension of concepts may vary, and that this would be particularly relevant to poor members. Deciding criteria for membership is a learning process which develops to maximize similarity within categories and dissimilarity between categories: i.e. under the rules of family resemblance (Rosch & Mervis, 1975). In this case, developmental variability is not due to conceptual change as such, but change in the criteria for acceptance and rejection. Thus, children's concepts should be basically the same as adult concepts, that is to say they should have the same central tendency, although their extensions will be different. That children's extensions are different from that of adults has been widely documented. At an early stage children's extensions are extremely wide e.g. "daddy" may be used to refer to every male. At a later stage children's extensions are too narrow, excluding bad examples from their categories (e.g. excluding "penguin" from BIRD). Stability between individuals in the same culture should be reasonably strong as, like the probabilistic views, concepts are formed in relation to a structured world. In particular, typical category examples should be highly correlated across subjects. Barsalou (1987) has reported that interpersonal and intrapersonal stability in graded structure is less high than was previously thought, although higher concordance is found for typical instances. He has also suggested that the Exemplar View can provide a reasonable account of this level of stability:

> Although this approach has not yet been oriented towards accounting for the instability of graded structure, it has potential for doing so. For example, between-subject instability could result from different individuals experiencing different exemplars over time... In general, what seems typical to a particular person on a particular occasion may reflect the exemplars that are readily available.

Barsalou, 1987 p. 132

2. Linguistic Function

This is the conceptual function for which the Exemplar View provides the weakest account. It has no adequate way of explaining semantic implication, synonymy or antonymy. A possible linguistic role for concepts seems, as yet, excluded from this holistic approach.

3. Simple Categorization

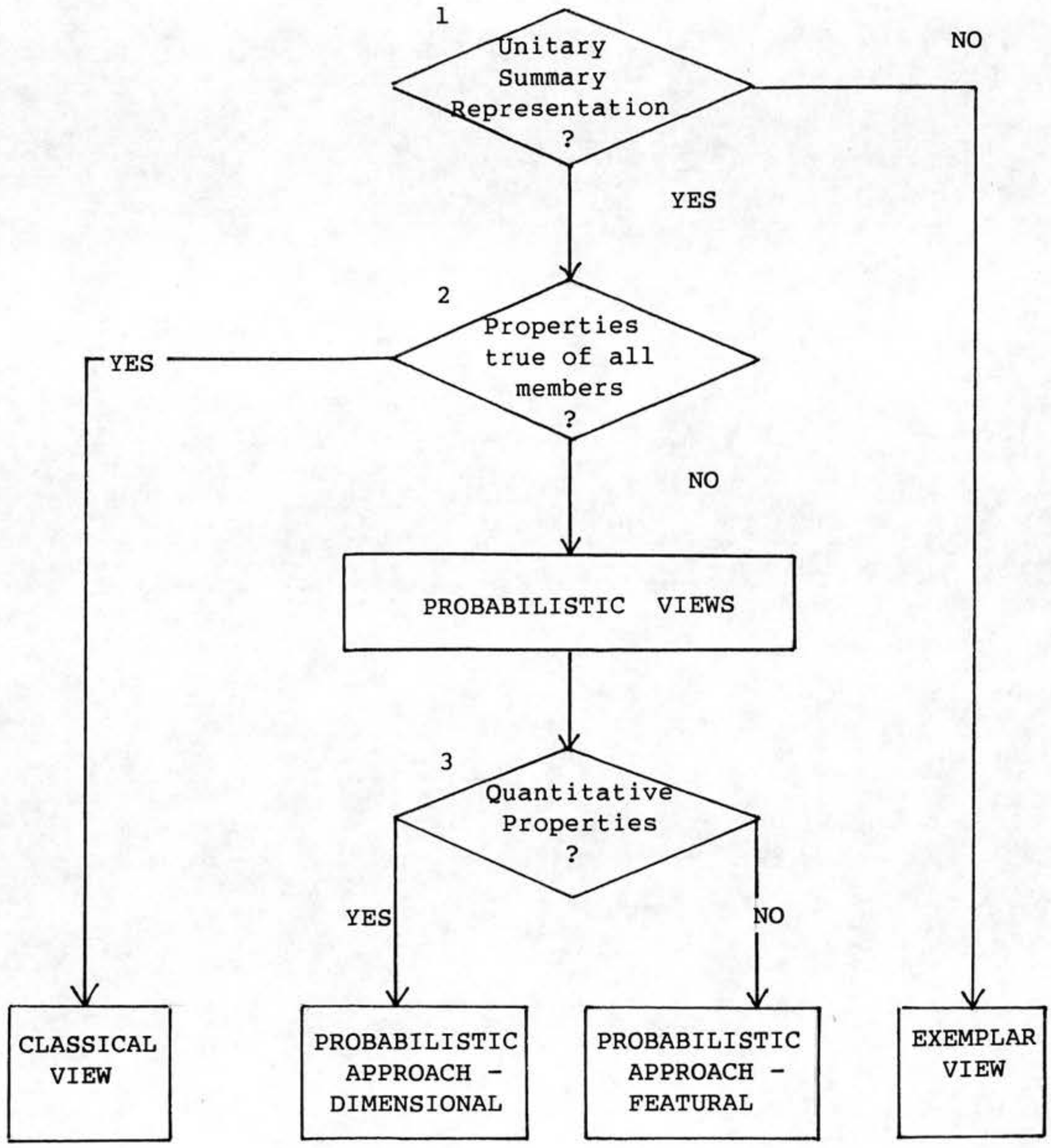
The Exemplar View provides an excellent account of typicality effects, referring to the fact that some examples will be closer to the prototype than others. In a strong version of the Exemplar View any notion of definitions is ruled out (Note 5). It simply makes no sense to say that FRUIT has APPLE and PEAR as its definitions; in this case all fruits would be apples and pears! Hence, this view finds it problematic to account for typicality effects in categories that are intuitively well-defined. Armstrong, Gleitman & Gleitman (1983) have shown that the categories odd and even number (which are intuitively defined e.g. "divisible by 2 leaving a remainder 1" and "divisible by 2 leaving no remainder" respectively) show typicality effects. It is difficult to see how a complete understanding of "odd" number could be grasped if this category is organized around comparison with the prototype "7".

4. Complex Categorization

Although I know of no-one who has actually proposed an uncompromised Exemplar account of complex categorization, it seems to me that a logical extension of the Exemplar View leads to some surprising predictions about complex concepts. Namely, that complex concepts must be independent from the similarity distributions of their constituents. Consider the example "pet fish", the constituent "pet" is, according to the Exemplar View, represented by a prototype, let it be DOG. The second constituent "fish" is similarly represented, let it be by the concept HADDOCK. By what possible means could these concepts combine to from a prototype for the complex concept PET FISH? Thus, because the Exemplar View, in its stronger form, does not include abstract information it must predict that complex concepts have whole new goodness of example distributions. Only this would allow the prototype for PET FISH to be something like "goldfish". Hence, it seems that the Exemplar View must abandon any notion of compositionality.

Given this state of affairs, it is now pertinent to ask what allowed the formation of the complex concept PET FISH at all? What is to stop the emergence of concepts like PET CHAIR or CHAIR FRUIT i.e. how are complex concepts to be constrained? An Exemplar theorist would argue that constraints are inherent in the environment. Thus, PET LOG is a meaningful conjunction because Yukka plant offshoots are sold with accompanying "names" and "characters" - some even sport stuck on plastic eyes for effect! So, conjunctions could be constrained by the non-existance of obvious referents in the environment. It seems that in addition to this, some general lexical knowledge is often brought to bear. For instance, even in noun-noun combinations a head-modifier relationship between the constituents is still recognized - see Chapter 4. Hence, even if the Exemplar View provides a good account of categorization procedure, its lack of reference to linguistic information casts doubt on its completeness as a theory of concepts.

FIGURE 1



A SUMMARY OF THE ASSUMPTIONS UNDERLYING THREE THEORIES OF CATEGORIZATION

Adapted from Smith & Medin, 1981

1.5. Towards A Theory of Concepts

Summary

In summary, it can be stated that there is, to date, no adequate theory of concepts. The three views summarized here each account for specific conceptual roles without reference to how Rey's other functions might be achieved. The Classical View fails to account for the categorization data, in particular typicality effects. On the other hand, it provides good explanations of linguistic function, and captures our intuition that words have definitions, by hypothesizing that concepts contain necessary and sufficient features. The Probabilistic approaches provide a good account of simple categorization phenomena, but cannot stretch to an explanation of complex concepts. The Exemplar View gives a holistic approach to categorization which agrees with the findings on basic level dominance. It also copes well with simple categorization, but provides no explanation of the linguistic function of concepts. The Exemplar View predicts that complex concepts will be represented separately from their constituents; but overall seems inadequate as a theory of concepts because it indicates no way of incorporating linguistic information. The differences in assumptions between these three approaches are outlined in Figure 1. Below, the criticism that none of these theories can be considered a theory of concepts will be examined, and an assessment made of what theoretical base does exist on which future research into the nature of concepts can be grounded. I will look again at the problem of complex concepts and suggest that they may provide an appropriate experimental field through which to expand our knowledge of concepts and how they operate. I will examine three main criticisms which have been raised against the work to date on concepts and categorization.

1. Epistemological / Metaphysical Distinction

Rey (1983) attacks the validity of much empirically based categorization theories, when he criticizes the theoretical account given of them by Smith & Medin (1981).

A similar argument has also been advanced more recently by Lakoff (1987). Rey asserts that the Probabilistic and Exemplar approaches are epistemological in nature, that is to say they address the means by which subjects classify things, not the conceptual representation of the things they classify. Rey does not question the value of knowing access patterns, he simply asserts that this knowledge does not constitute a theory of concepts.

If Rey's point is that observing behaviour (categorization) is not synonymous with observing cognition then what he says is accurate, if not entirely new. (It would be a shock indeed if we had not learned the lessons of behaviourism). Rey's criticism is, however, a timely reminder not to let our rhetoric or thought patterns slip into that error. On the other hand, Rey seems to completely ignore the fact that there are good reasons to *hypothesize* that changes underlying representations may have some bearing on categorization decisions. Prototype theory (Probabilistic and Exemplar theories) was originally developed as a possible explanation of the data of typicality effects. A theory is, after all, a proposed explanation for a phenomenon. Moreover, prototype theory generates predictions, and can therefore be falsified. Thus, I do not think that classification data can be rejected out of hand as irrelevant to theories of representation. It is particularly difficult to discount Rosch's findings that priming with a category name facilitates "same" judgements for identical typical category pairs, but hinders "same" judgements for identical atypical pairs. It would be very difficult to provide an alternative explanation for this finding without recourse to underlying representations. In the final analysis, Rey highlights the fact that an adequate theory of concepts has not yet been developed. It is my reply that the study of conceptual usage, particularly complex categorization decisions, is the most promising road towards one.

2. The Problem with Levels

Smith & Medin (1981) have drawn attention to an important problem facing anyone trying to develop a theory of concepts: that is, at what level does conceptual representation occur? The confusability of "E" and "F" in letter list experiments may be explained by referring to shared line features, for instance. At another level "E" itself might be a feature used to explain the confusability of the words "BED" and "RED". At yet another level "BED" might be one of the features used to explain the confusability of the sentences "The boy went to bed" and "The boy went to sleep". And so on. Smith & Medin simply assert that there are empirical grounds for distinguishing between features and concepts: thus they argue that it is empirically valid to conclude that "E" is a feature in the "BED", "RED" example. Smith & Medin conclude that: "what is a concept at level 'n' is best thought of as a feature at level 'n+1'".

Smith & Medin's treatment of this issue seems rather flippant. They are, in effect, drawing attention to the fact that current theories of concepts imply huge conceptual networks for different conceptual levels. If this is true, it may also be true that conceptual representation is concentrated in a "basic" most salient level, somewhere in the middle of such a taxonomy. This idea is not a popular one, and the notion that there exist conceptual primitives, features that are themselves indivisible, is preferred. These problems are as yet unanswered, but a complete theory of concepts will have to take them into account.

3. Well-Defined Categories

Armstrong et al (1983) have shown that typicality effects occur in categories that are well-defined. Thus, they argue that finding typicality effects in natural categories where defining features are unknown, is not evidence to suggest that they do not exist. While this is true it is important to note that the opposite is just as true. That is, even if definitions can be provided for a category, that does not mean that people actually represent the category in terms of that definition.

It is being increasingly felt that the Classical View may still have a lot to offer a theory of concepts. Some researchers are combining the notion of defining features and characteristic features in concepts to form fuzzy concepts with defining "cores". In particular, the Classical View may be pertinent to the understanding of verbal concepts (Miller & Johnson-Laird, 1976). Again, the conceptual core may be used to explain the linguistic function of concepts (Armstrong et al, 1983), whereas characteristic features may be more fully involved in the identification procedure of items. For example, Barsalou (1982) has suggested that it is only the characteristic features of concepts that are subject to contextual change. The core aspect of concepts may also play a particularly vital role in placing constraints on possible concepts. How these postulated constraints could operate in the formation of complex concepts, however, is unclear. Recently, James Hampton has proposed an account of how complex concepts might be formed when the core and characteristic functions of the two constituent concepts combine (Hampton, 1987). He makes use of an "inheritance of attributes" rule where usually a complex concept is formed by the union of the *intensional* representations of the constituent concepts. (This emphasis on intensional combination avoids some of the problems of the Classical View - see section 1:2). Hampton also provides a number of alternative approaches when the inheritance of attributes from both constituents conflict. While this theory is hopeful in that it addresses some of the complexities of conceptual combination, it is open to the criticisms levelled against all featural approaches. Namely, that this approach is untestable while there is no way of knowing the particular features which make up the concepts involved - see section 4:6.

In conclusion, theories of concepts which have developed in response to categorization phenomena are a valid step towards a complete conceptual theory. If the problem of levels is taken seriously, what we have learned about natural kind taxonomies may also be pertinent to the mental organization of concepts. Finally, the Classical View may still have a role to play in meaning; a role that may develop as different kinds of words and categories are studied.

1.6. Concept Combination

It has been argued in this Chapter that current theories of concepts do not appear to provide an adequate account of how concepts combine. The Classical View predicts that concepts will combine according to a Boolean model. That is, a complex concept will have only those attributes which belong to both constituent concepts. As discussed earlier (see Section 1.2:4) these criteria do not hold for many combination cases. A posible solution to this problem is to propose that concepts are combined intensionally, rather than extensionally. In this case, the extensional set for the complex concept need not necessarily have a predictable relationship with the extensional sets of the constituent concepts. The problem with this solution is that an empirical demonstration of intensional combination seems impossible.

The Probabilistic views attempt to incorporate graded membership into a model of conceptual combination, suggesting that two "fuzzy sets" combine to form a third. This view fails to account for the fact that it is not always the most typical aspects of the constituent concept that are included in the conjunction. In contrast, it seems that the particular concepts combined, highlight in some way certain features, making them more salient, in that context. Thus, a Fuzzy Set model fails to capture the *interactive* nature of conceptual combination (see Section 1.3:4).

The third theory considered was that of the Exemplar view, which in its strongest form suggested that concepts do not combine (Section 1.4:4). Rather, new concepts are continually being learned which might then be associated with one or more words to which a meaning is already ascribed. This theory has the practical problem of being non-productive when human processing capacities are limited. In addition, it has the theoretical problem of providing no constraints on concept formation. On the other hand, it may be true that conceptual combination can lead to the permenant formation of new concepts. That is, there seems to exist a stage where a conceptual combination becomes so common that it has an identity distinct from that of its constituents (e.g. "ashtray" is not usually thought of as a conjunction of "ash" and "tray"). Such combinations subsequently develop independently of their constituent concepts, and after some time only their etymologies may show they have arisen from conceptual combination. It should be expected that in a language at any given time there will exist many words and combinations in various stages of this "lexicalization" process. Thus, any given word pair may represent a different relationship with the constituent concepts of the combination.

Despite these difficulties, conceptual combination is a crucial issue for conceptual theorists. This is because meeting concepts in different sorts of contexts and in different sorts of combinations is the rule of language, rather than the exception. Complex concepts can be construed as a special sort of context effect. When a complex concept is presented in isolation each word can be said to act as a context for the other. Thus, when presented with the complex concept "pet fish", "pet" provides a context in which to interpret "fish", and "fish" provides a context in which to interpret "pet". Further, Barsalou (1987) has argued that context effects in general can be described and understood as the results of routine conceptual combination. Thus, complex concepts need not even be linguistically marked by a word pairing like "pet fish", but may be contained within a sentence: "His fish made a good pet".

In order to investigate the processes of contextual change and conceptual combination, processes which may be intrinsically linked, I will make use of what is already known about the structure of noun categories. I will study how category structure changes in different contexts and also how noun categories combine. In this way I hope to be able to discover principles of conceptual change which are applicable to conceptual representations in general. That is not to say that all concepts are represented in the same way, but rather that discovering something of how everyday natural kind noun categories function, may provide insights into other concepts.

CHAPTER 2

CONTEXT EFFECTS ON CATEGORY CONCEPTS

2.1. Context and Similarity to a Prototype

In Chapter 1 several skeleton models of conceptual theories were studied. The only application of those theories considered were one-off simple or complex categorization judgements. This was primarily because, as difficulties in explaining complex categorization may have indicated, none of these theories seem able to cope with the knowledge rich environment in which people normally categorize. When categorization is investigated in such an environment, it becomes apparent that all categorization decisions are very context sensitive. This means that, for example, while in the context of international competitions, "chess" might be judged to be a member of the category "sport", it might be judged to be a non-member in the context of a school P.E. lesson. Context does not only affect the likelihood of atypical instances being included in the category, it also affects typical instances. For example, the prototypical referent of "fruit" is "apple" or "orange" but the prototypical referent of "fruit" in the context of "fruit for making jam" is "raspberry", "strawberry" or "blackcurrant".

These contextual changes are relevant to the stability function of concepts. The Classical View asserts that all contextual changes in goodness of example (GOE), can be explained with reference to the procedural aspects of classification. Or, that contextual effects can be explained in terms of changes in the characteristic features of concepts. Thus, it is argued that context does not alter the metaphysical aspect of concepts nor undermine conceptual stability *per se*. If this is so, how is it that in some contexts a person will consider an item to be a category member and in other contexts a non-member? If this does not challenge conceptual stability it is unclear what could. It is also unclear how a Classicalist could explain the occurrence of contradictory membership judgements in the same individual at the same time. A possible explanation would be the assertion that the characteristic features accessed in a given context have as much of a role in determining the categorization decision as core features (such a position seems to be proposed by Barsalou e.g. Barsalou, 1983). In this case, the Classical and prototype views seem difficult to distinguish, since the Classical View is effectively abandoning the idea that membership is calculated on properties that are true of all members. (Barsalou, 1982) has gone further than this, describing cases where context independent (core) features can be "inhibited").

For the prototype views (Probabilistic and Exemplar Views), the problem of stability is not so difficult to explain. The prototype views have always placed the onus of accounting for stability outside the conceptual domain and rather, for example, in the structure of the environment. In the case of context effects, the environment has in some sense altered, and thus concepts reflect that change. As the environment only changes in largely predictable and limited ways, boundaries are set on conceptual change too, thus solving problems of conceptual constraints. This argument, that the environment constrains concepts, while *prima facie* true, is oversimplistic. As Quine has demonstrated in his famous "translation argument" (Quine, 1960) any perceived event is highly complicated, and it remains a mystery how people correctly select salient episodes as relevant. For example, the scene of a dog rushing into a room to greet his master, carrying a toy in its mouth, provides a context which could be said to constrain the the category concept "TOY". How the relevant constraining information is selected from this complex environmental scene is unknown. What seems certain, however, is that information will be selected only after the scene has been "simplified" and interpreted through our current conceptual structures (Kahneman & Tversky, 1973). Thus, to argue for environmental constraints entails a certain circularity since the perception of that environment is itself constrained by the concepts held.

Another problem for prototype theories is in explaining how contextual changes are organized and represented. Given that we do have some knowledge of word meaning out of context, and that contextually appropriate meanings can be quickly supplied, there is clearly a great deal of information available. The simple representational idea of computing similarity to a prototype may have to be altered to deal with contextual change.

The questions of context and conceptual representation seem inextricably bound together. In this chapter I will consider the extent to which context effects dictate a change in prototype theory. In section 2.2 possible explanations for contextual change in GOE distributions are discussed, and a study of GOE in sentence contexts reported. The implications of this study are discussed with reference to the issue of conceptual representation. Section 2.3 concludes the chapter with a discussion of the issues raised.

2.2. Experiment 1

The Effect of Context on Reported Goodness

of Example in 11 Noun Categories

One of the biggest challenges to the adequacy of prototype theories as theories of concepts is the well-attested fact that the perceived goodness of example of an instance changes with context. What is not clear is the extent of variation in the prototypicality distributions for the same category name in different contexts. It is possible, for instance, that it is only the *prototype* of the category which changes in context. As GOE is calculated with respect to how similar an instance is to the prototype, then it should be possible to change the prototype of the category while maintaining the stability of the GOE distribution. If this "refocusing" hypothesis were accurate it would predict that contexts yielding the same prototype would also yield the same GOE distributions: as GOE is, by definition, computed relative to the prototype. The alternative hypothesis is that the whole GOE distribution is generated anew in different contexts. In this case, contexts having the same prototype need not have the same GOE distribution, as the whole distribution is computed relative to the context.

Roth & Shoben (1983) report a number of experiments which address the question of how context alters the GOE distribution. Firstly, they rule out the refocusing explanation of context effects by showing that categories which have the same prototype do not necessarily have the same GOE structure. Roth & Shoben devised pairs of sentences which suggested the same prototypes but different second-best examples. One context sentence suggested a typical instance as second-best example, and the other an atypical instance as second-best example. As they had predicted Roth & Shoben found that subjects would rate an instance as more typical than another instance in one context, and less typical than that instance in another. For example, subjects were presented with the sentences: (1) During the midmorning break the two secretaries gossiped as they drank the beverage; and (2) Before starting his day the truck driver had the beverage and a donut at the truck shop. Both sentences suggest "coffee" as the prototypical referent. Subjects rate "tea" as more typical than "milk" for sentence (1), but, for sentence (2), "milk" (a normally atypical example of "beverage") is rated as more typical than "tea". Thus Roth & Shoben found that even when sentences suggest the same prototype they do not necessarily have the same GOE distributions. In this way they were able to rule out the refocusing hypothesis as an explanation of context effects.

Their findings also contradict an amended form of the refocusing hypothesis which suggests that context alters the category prototype and also imposes a selection restriction on a static GOE distribution. For example, the GOE norms in Rosch (1975a) for the category "fruit" give the most typical exemplars as: "orange, apple, banana, peach, pear...". The selection restriction hypothesis predicts that subjects would report the same GOE distribution as that generated in a 'neutral' context excluding misfitting exemplars. (Or, if it was appropriate in the context, subjects would alter their prototype for the category before applying the selection restriction). So for a context such as "The boy had to peel the fruit before he could eat it" the selection restriction hypothesis predicts that subjects should give "orange and banana" as the most typical exemplars (simply omitting "apple", "peach" and "pear"). As Roth & Shoben have shown, however, it is not merely the case that some contexts cause exemplars to be excluded, exemplars also change places in the typicality distribution. In the "beverage" example, "tea" and "milk" were both possible referents of the two context sentences, but their representativeness was reversed from sentence 1 to sentence 2. Thus Roth & Shoben have demonstrated that it is not only the focus, or the prototype, of a category which alters with context, but the GOE distribution itself.

However, Roth & Shoben only studied the effect of context on category prototypes and one typical and atypical exemplar. Could it be the case that the GOE distribution is, in fact, relatively stable across contexts with only a few context relevant exemplars shifting in representativeness? In Experiment 1 the effect of context on the entire GOE distribution will be considered.

In a recent paper Barsalou (1987) has suggested that a possible explanation of the context effects demonstrated by Roth & Shoben is that a complex concept is being formed during processing. Barsalou suggests it is this fact that makes the distribution of the simple concept appear so unstable. So, for example, in understanding sentences like "Wendy loved to ride the animal" and "Wendy loved to milk the animal" subjects could be forming complex concepts of ride+animal and milk+animal, respectively. Since the complex concepts formed in these instances are different it is hardly surprising that the typicality of exemplars is also different.

If Barsalou's suggestion is correct and complex concepts are routinely made when processing sentences of this type then a clear prediction follows. It should make no difference to typicality ratings if key context words are presented as complex concepts, or are presented as modifiers in other parts of the context sentence, providing that the meaning of the two sentences does not change. Thus, if Barsalou's suggestion is correct, the two sentences: (1) "The teenager had already acquired some small articles of furniture of modern design", and (2) "The teenager had already acquired some small articles of modern furniture of her own", would be understood by forming the complex concept "modern furniture" in each case. Although it would be expected that the linguistic form "modern furniture" might make it easier to form the complex concept, it should not actually make any difference to the typicality ratings given by subjects. This is because subjects will have to have understood the sentence (and hence formed the concept) before judging which items are typical of it.

In order to test this prediction context sentences for Experiment 1 were of three forms. The first, 'simple' form was to act as a baseline control and included the category name without the context modifying word e.g. "The teenager had already acquired some small articles of furniture of her own". The second 'complicated' form included a modifying word which was in a separate part of the sentence from the category name (as 1 above). In the third 'complex' form the modifying word and the category name formed a complex concept (as 2 above).

The reasons for this study are threefold: firstly, to allow the scale of contextual changes on typicality to be assessed; secondly, to investigate the effects of two linguistic forms on typicality; and thirdly, to provide a source of accurate typicality norms for specific context situations. This last reason enables experimental performance in specific contexts to be compared with known typicality ratings *for that context*, as well as typicality ratings obtained in a 'neutral' context. The study was carried out in a similar fashion to the collection of GOE norms by Rosch (Rosch, 1975a). Distributions were collected for several categories by embedding the category names in a number of sentences.

METHOD

Stimuli

The categories for which ratings were obtained were chosen from the concrete noun categories documented by Rosch (1975a) and Uyeda & Mandler (1980). All those categories used by Rosch, save the category "TOY" with which she experienced problems, were selected. In addition, two categories from Uyeda & Mandler were chosen. This resulted in a total of 11 categories, namely: FURNITURE, FRUIT, VEHICLE, WEAPON, VEGETABLE, TOOL, BIRD, SPORT, CLOTHING, INSTRUMENT and CLOTH.

Participants

Participants were 88 young people who were either psychology undergraduates or youth hostellers. All participants were native English speakers. The sexes were approximately equally represented in the sample. All those who took part in the study did so voluntarily.

Materials

Each category name was incorporated in a sentence such that the sentence intuitively denoted either a typical or an atypical member of the category. Sentences were largely adaptations of those used by Potter & Faulconer (1979). Additional sentences were adapted from common books for older children. The sentences had the special characteristic of being renderable in three forms. In the most complex form the noun is qualified by a prenominal adjective; e.g. "The woman screamed "Help!" as she stared at the *bloody weapon* in his hand". The second, "complicated", form contains a noun qualified by a adjective in a different clause or sub-clause; e.g. "The woman screamed "It's *bloody!*" as she stared at the *weapon* in his hand". The simple form omits the adjective from the sentence altogether; e.g. "The woman screamed "Help!" as she stared at the *weapon* in his hand". These formats are particularly useful for investigating the difference between adjectival qualification and noun-noun combination, as they provide appropriate semantic and linguistic controls.

Three sentence lists were designed such that each list contained all three forms of sentence but only one form of each sentence. There were 24 sentences in each list. (After the start of the experiment an additional 15 sentences were included in each list in order to increase the proportion of atypical sentences, which was found to be less than half). The sentences were typed on A4 sheets, spaced so that subjects could write responses underneath each sentence. There were approximately 11 sentences per page.

Procedure

Participants were informed that the purpose of the experiment was to obtain general normative lists of the category members people thought of when they read the sentences. Participants were given a list of sentences (lists were assigned randomly) and asked to write down the category member they thought best fit the context of the sentence. They were given an example of how they might complete the task. Participants were also asked to write down the exemplar they felt was second-best, in order to increase the variance in responses. The participants task was to *produce* an instance, rather than order a set of given instances, as it was reasoned that the former, while less usual in experiments, would better reflect many experimental situations which utilize implicit category decisions. Once the task had been explained, the participants completed the questionnaires in their own time. Each sentence contained only one category name, and the category was always central to the meaning of the sentence. The category names were not marked in the sentences as it was reasoned that this might distract from a holistic reading of the sentences. Participants found it obvious which word was the category name.

An average of 23 people rated each form of a sentence. As 15 sentences had been added after the start of the experiment 20 of the 88 participants rated only these 15 sentences, in order to balance the number of people rating each sentence. As before, each participant rated all sentence forms but only one form of each sentence. In no case did the ratings of this additional group alter the prototype for the sentence.

Ratings

Exemplars listed by the participants were collated into a production frequency list for each sentence form. The prototypical exemplar was deemed to be the exemplar cited by most people as the best exemplar. However, in the case where three forms of a sentence shared an exemplar in their three most typical exemplars of the category, that exemplar was deemed to be the prototype for all three forms. This rule was adopted so that obviously similar distributions would not be represented as suggesting different prototypes. If there was a draw between two exemplars, the prototype was deemed to be that exemplar which was most often listed as second-best choice as well. The typicality rating for the sentence was the rating given to the prototype in the wide listings of Uyeda & Mandler. In the case where the exemplar was not included in these norms, its position in them was estimated using the longer listings of Rosch. If the word did not exist in either source it was given a rating equal to the worst instance cited in Uyeda and Mandler's list for that category.

A sentence was given the label "Typical" or "Atypical" according to whether its prototype fell above or below the typicality rating of 2.5. (This is approximately half way between the two extremes listed in Uyeda & Mandler's norms). In the case where the typicality rating of a sentence lay between 2.3 and 2.7 inclusive, it was given a label "Typical" or "Atypical" only if the two next most frequently cited exemplars also fell on the same side of 2.5, otherwise it was deemed a borderline case unsuitable for use in future experiments.

Eight sentences were excluded from the study as participants had misunderstood the modifier, or thought the sentences odd, or, in the case of "instrument", had found the category too unconstrained.

Thirty-one sentences, each of three forms, were considered appropriate for further investigations. The average typicality rating for Typical sentences was 1.5 and for Atypical sentences was 4.7. The average sentence length was 13 words. The complete listings of the sentences can be found in Appendix 1.

Change in Linguistic Structure

The normative distributions for 31 context sentences were examined to see if the linguistic form of a sentence has an effect on the typicality of instances, when the sense of the sentence remains the same. In particular, are distributions obtained for categories with prenominal modifiers (in which the reader is thought to have formed a complex concept) the same as distributions obtained for categories with modifiers in separate clauses? If so, this could imply that complex concepts are routinely constructed when understanding sentences, and that context effects are explicable in terms of a person forming different complex concepts in different contexts.

In order to address this question prototypes for complicated and complex sentence forms were compared for each sentence. (It was not appropriate to compare the whole distributions obtained for each sentence form. This was because many of the less typical items had been named by only 1 subject, and were thus unreliable). For 26 of the 31 sentences the prototypes were identical and the distributions similar. 4 sentences had very similar prototypes e.g. the second-best exemplar in one form, might be the best exemplar in the other. For only one sentence were the prototypes for the two sentence forms different. This was for the sentence: "The retriever brought back the (scavenger) bird that the hunter had seen (scavenging)". The prototype for the complicated form is "pheasant" but for the complex form is "vulture". An explanation for why this case is different from the others could be that for most people "scavenger bird" is a single concept, that is to say, it has become lexicalized. Thus, participants may already have a stored meaning for "scavenger bird" in a way they do not for "modern furniture".

The findings of Experiment 1 are consistent with Barsalou's idea that people may be forming complex concepts when encountering categories in context. The fact that changing the linguistic form of sentences in the way described does not affect the typicality of exemplars gives credence to the idea that complex concepts might be a simplified case of context effects. The different result in the case of "scavenger bird" indicates that very different effects should be expected from common complex concepts, the meanings of which people may already have stored - see section 4:6 for further discussion of this process.

Restructuring, Refocusing or Restriction?

The normative distributions obtained in Experiment 1 illustrate that context does alter the perceived typicality of instances of a category. Many context sentences made normally atypical exemplars the prototypical referent. Even in cases where sentences were judged to be typical, the instances included in the distributions, and their ordering varied for different sentences in this study, and from the norms documented by Uyeda & Mandler and Rosch. For instance, the prototype for "weapon" in the sentence "The woman screamed "Help! / It's bloody!" as she stared at the (bloody) weapon in his hand" was "knife". However, the prototype for "weapon" in the sentence "Though Sam had it with him (in a harness) he hoped that the (harnessed) weapon wouldn't be necessary" was "gun".

These results cannot be explained by the refocusing hypothesis as (like Roth & Shoben) sentences suggesting the same prototypes were not equally suggestive of other instances. For example the sentence: "Charles carefully disassembled and cleaned the instrument he had found" suggested "trumpet" as prototypical, "flute" as quite typical and "trombone" as less typical. The simple form of this sentence: "Charles carefully cleaned the instrument he had found" was more suggestive of "flute" than "trumpet", and "trombone" was not mentioned in this distribution. Thus, even slight changes in meaning can alter the GOE distribution and those alterations cannot be explained in terms of changing the focus of a fixed distribution.

The fact that the typicality of "flute" and "trumpet" reversed in different contexts also rules out the selection restriction hypothesis. Items are not simply being omitted, their typicality is changing and they are emerging elsewhere in the distribution. This restructuring cannot simply be explained in terms of a slight instability of typicality judgements over time, or between people, because in some contexts items that are normally atypical category referents become prototypical category referents. For example the sentence: "He won the fairground fruit in a competition" most strongly suggests the atypical fruit "coconut" over "apple" and "orange".

The results of this study confirm that the total GOE distribution is altered by context, as was suggested by the results of Roth & Shoben (1983). The question which now has to be answered is whether these changes reflect conceptual changes. Could the change in the typicality distribution be a result of inference processes, or combination processes, that occur after the concept has been accessed? Or does the change in typicality imply that there are no stable prototype structures to concepts? These questions emphasize the importance of assessing whether contextual restructuring occurs at a conceptual level on the one hand, or an inferential level on the other. As these norms provide information about which category instances are contextually appropriate in a number of different contexts it is now possible to compare the explanatory power of context-dependent prototypes with those obtained when context is unspecified.

2.3. Is Similarity Enough?

Experiment 1 served to clarify the fact that the typicality structure of a category is radically altered by changing the context in which the category name occurs. Changes in the linguistic structure of a sentence (whether the noun was qualified prenominally, or by a word in a different clause) did not cause a change in typicality ratings. This is consonant with the view that context effects may be the result of the combination of the key concepts in a sentence. A further study (see Appendix 2) sought to discover whether contextual changes are a result of changes at a conceptual level or a process of post-access inference. Neither Rosch's prototypes or the context-appropriate prototypes of Experiment 1 were able to predict performance on a Stroop task in this experiment. In this case, the wide range of within and between subject variation suggests that context has wider ranging effects than can be accounted for in terms of changes in a goodness of example distribution. In Experiment 1, the goodness of example distribution was found to be completely altered by context, disallowing a small amendment to save the prototype views. If similarity to a prototype (even a context-appropriate prototype) is not sufficient to explain context effects then prototype theory cannot be an adequate theory of concepts.

There is a growing consensus that similarity to a prototype cannot adequately account for conceptual structure. The notion that it is similarity which provides conceptual coherence is appealing because it seems intuitively obvious that things are grouped together in categories because they are fundamentally the same. This intuition is supported to some extent by the perceptual similarity of basic level categories. Rosch, Mervis, Gray, Johnson & Boyes-Braem (1976) found that suitably scaled and oriented drawings of members of basic level categories (e.g. different sorts of chair) when superimposed had a high degree of overlap, and a discernable (chair) shape. Neisser (1987) has pointed out that basic level categories also have a high level of functional similarity. We "sit on" rocking chairs, sag bags, kitchen chairs and umpire chairs; we "put things (we want to interact with) on" coffee tables, kitchen tables, dining tables, pasting tables. Similarity has thus provided a suitable explanation for the salience of the basic level, and this success has made plausible the theory that it is similarity that gives concepts, in general, coherence.

Many authors are now pointing out the limitations of similarity as a complete account of conceptual coherence. Firstly, similarity weightings cannot represent relations between concepts. For example, taxonomic relations between concepts cannot be captured by similarity information. Secondly, perceptual similarity does not always aid classification, as Robert McCauley (1987) has pointed out: "whales are not fish, spiders are not insects, and palm trees are grasses". In these cases perceptual similarity has to be ignored in order to make a correct classification. However, the very fact that people find these cases difficult to classify shows that perceptual similarity plays some role in categorization. Correct classification of these items occurs when people pay attention to less salient attributes. As a similarity model can provide no explanation for favouring an item's less salient attributes, it cannot explain conceptual structure. It is an important point that having a well-defined category does not solve the problem of coherence either. A category might be defined as: "things that weigh more than a ton which are not blue", this is well-defined, but it is not coherent.

The crucial question is how do people know which attributes are relevant in a given categorization scenario? Any two items can be arbitrarily similar or dissimilar depending on which attributes are compared. For instance, a person might say that "whale" is a member of the category "fish" if they were concentrating on attributes such as "lives in the sea", "eats plankton", and its perceptual similarity to other items e.g. "is like a shark". If a person concentrated on other attributes such as "is warm blooded", "breathes air" and "has a blowhole", "whale" would be categorized as more similar to mammals than to fish. Thus similarity is not an explanation of categorization, but a result of which attributes are chosen for comparison. This point is reiterated in Barsalou's (1983) finding that ad hoc categories have a similarity gradient although there is no *a priori* reason to think the category members are more similar to each other than to non-members. For example items such as "jewellery, coat, photograph, child, teddy bear, bank book" do not seem to form a cohesive category until one is given the category name "Things to take from a burning house". This information clarifies which attributes are important in this context.

Thus it is not overall similarity between objects that is the important factor in category structure, but similarity computed over the relevant attributes. Which attributes are relevant for a given category will change with the context in which it is used, and with the purpose it serves. What holds a category together is the underlying processes responsible for selecting the appropriate attributes. Thus, the notion of similarity to a prototype is no longer doing the work in explaining category cohesion.

Similarity may, in fact, be a by-product of categorization. Items may appear to be similar simply by virtue of the fact that they have been classified as members of a category. Thus, similarity may not explain why an item is classified as a member of a category, it may be the result of such a classification. The fact that category membership does lead to an increase in perceived similarity within groups, and perceived dissimilarity between groups, is a well known effect. For example, if subjects are presented with a series of coins each increasing in size by a standard amount, subjects will judge the difference between any two coins to be greater if those coins are on the boundary of an arbitrarily labelled group. Thus, simply drawing a line between the 4th and 5th coins, and labelling the first four coins "A" and the second four "B", increases the perceived similarity of the coins within groups "A" and "B", and increases the perceived difference between "A" and "B" (Tajfel & Wilkes, 1963).

Medin & Watenmaker (1987) have also argued that similarity may be a by-product of conceptual coherence rather than its determinant. They use the example that winning basketball teams have in common the fact that they all score more points than their opponents. In order to explain why the teams score more points one must turn to more basic principles. In the same way as scoring more points at basketball is a by-product of being a winning team, Medin & Watenmaker argue that similarity may be a by-product of category membership.

If similarity is not an adequate solution to category coherence what can replace it? How do people select which attributes are relevant in a given situation? In the next chapter I report an investigation of the integration of context information during sentence reading. I address the question of how our understanding of a category is affected by the context which precedes it. By answering the question of when people correctly select context relevant attributes when processing categories in context it may be possible to gain a clearer understanding of what makes concepts coherent.

CHAPTER 3

THE MEANING OF NOUN PHRASES

3.1. Category Coherence

Conceptual coherence is thought to be important for two reasons. Firstly, it is thought that coherence holds a clue to meaning. It is argued that items classified as members of a category "belong together" independently of that classification. Evidence for this view is provided by the fact that: (i) children will spontaneously group these items together when asked to sort items into piles, and (ii) category members often seem to be constrained ecologically, for example, they may afford the same properties. Thus, it is thought that if the underlying representational commonalities could be worked out (definitionalists), or the appropriate features and their weightings could be discovered (prototype theorists), then the meaning of category words would be clear. The second reason that coherence is thought to be important is that it can provide constraints on reference. It is argued that even if the meaning of category words is not fully known, and cannot be worked out, coherence at least limits the meaning of the category name. It is firmly believed that words must have a limited application, otherwise, it is assumed, effective communication would become impossible.

In order to explain how people maintain coherent categories across many different contexts it is important to investigate, in detail, the changes and modifications people make as a result of alterations in context. For instance, if changes as a result of context are made after a referent has been retrieved then it might be expected that the root of coherence is conceptual. This is because the same concept is accessed on all occasions, and any modifications are made on the same, underlying, basic units. The fact that concepts which occur in many contexts seem coherent can be explained by the fact that the same concept is at the root of all more developed meanings. Hence, it could be argued that concepts must provide "rules", or means by which such modifications can be constrained.

The converse of this argument would apply if it were found that context influences which concept is accessed. In this case, what provides category coherence across different contexts cannot be the conceptual representation as such. The solution to coherence must be provided by relations; that is, in terms of what unites different representations accessed in different contexts. Now the burden of explaining coherence is removed from a conceptual level to a level of relations between concepts.

The question I would like to raise is whether category coherence necessarily constrains reference. It has already been noted that many arbitrary groups of items can form coherent "ad hoc categories". This is because coherence is not a function of the items being classified. It seems that there may be nothing intrinsic to the items that makes the category coherent, or gives meaning to the word that names them. Both meaning and coherence are imposed on categories by people (Medin & Watenmaker, 1987). This leads me to the conclusion that categories are only constrained by people's ability to relate things together. I am not arguing that the environment can place no constraints on meaning, but that those constraints operate on people, not on words.

Barsalou's study of ad hoc categorization shows that coherence does not exist independently of categorization. It is only once a categorization has been made that coherence emerges. This is an important point as it leads to the conclusion that meaning and coherence do not exist independently of people: it is something imposed on the environment. Thus it can be concluded that meaning is primarily constrained from within. This means that a solution to the problem of coherence cannot be found extensionally, it must be found in the intensions of words, or theories concerning them.

The question that remains is whether *concepts* have an intensional structure which can provide conceptual coherence. One possible means to answer this question is to assess whether context exerts an influence at the conceptual level.

3.2. Experiment 2

On-line Adjectival Modification of Nouns

Potter & Faulconer (1979) designed an experimental paradigm capable of teasing apart whether context has an effect at the encoding or retrieval stage. They compared recognition latencies for pictures which were typical of a noun with pictures modified to fit the sentence context. For example, the time taken to recognize that a picture of a house matched an object in the sentence: "It was already getting late when the man first saw the burning house ahead of him" was compared with the time taken to recognize that a picture of a burning house matched the sentence. It was argued that the picture of a house (unmodified) is more typical than the picture of a house on fire (modified), this was supported by the short decision latencies for typical probes in sentences without a prenominal adjective. Thus, subjects should be able to identify typical probes more rapidly than modified probes if they are accessing the meaning of "house" without reference to previous context. If, on the contrary, subjects find it easier to recognize a picture of a "burning house" then context is already having an effect at the moment the picture test is given.

Potter & Faulconer (1979) compared responses to probes presented immediately after the target noun, with those presented at the end of the sentence. They found that for sentences in which the noun was modified by a prenominal adjective, subjects responded more quickly to modified pictures than typical pictures. Moreover, this result did not change for probes presented immediately after the target noun. This supports the proposition that word meaning is retrieved with reference to the preceding context, or constructed online in a contextually appropriate way.

One possible problem with the interpretation of this experiment is that it is unclear that subjects are matching the picture with a meaning derived from the sentence in each case. When presented with a picture probe "burning house" subjects may be able to make a quick match with the overall meaning of the sentence. When presented with the probe "house" subjects may have an initial mismatch with the meaning of the sentence as a whole, but use the strategy of echoing the sentence, thus finding a phonetic match with "house". This is possible because the phonetic recoding of pictures of common items is automatic (Snodgras & McClure, 1975; Paivio, 1971). The possibility of the introduction of this kind of strategy is sufficient to explain the longer response times for typical probes in the context of a sentence.

This potential problem can be circumvented by asking subjects to make a match between phonetically dissimilar probes and nouns. This can be achieved by asking subjects to make category judgements. The parallel of Potter & Faulconer's task would involve subjects judging whether a typical instance, or a modified typical instance, matched a category noun in a context sentence. For example, subjects presented with the sentence "'You found it' said Andy, noticing the broken tool in the workshop" would be probed by a typical (unmodified) picture "hammer", or a modified picture "broken hammer". The task would be to judge whether "hammer" matched an object in the sentence, in this case "tool". In this experiment a phonetic strategy could not work, and any matching must now be in terms of semantics.

The question addressed by Experiment 2 is how Potter & Faulconer's result transfers to decisions about category reference. When subjects are presented with a category name in a sentence context suggesting a typical exemplar do subjects find it easier to recognize a picture of a typical exemplar, or of an appropriately modified exemplar? This paradigm allows the assessment of whether category modification is immediate and automatic (i.e. outwith the requirements of the task) by comparing the response latencies to unmodified and modified probes when they are presented immediately after the occurance of the noun, or delayed until the end of the sentence.

Design

The design and procedure followed Potter & Faulconer (1979) save that only prenominal adjectival modification was studied. The design was mixed, the independent variables were probe type and position of probe. Probe type was the within subjects factor and had three levels: pictures that matched an object in the sentence but not the meaning of the sentence as a whole, pictures that matched both an object in the sentence and the meaning of the sentence as a whole, and filler items where the picture and sentence were unrelated. Position of the probe was the between subjects factor and had two levels: immediate - immediately after the mention of the object in the sentence, and delayed - after the end of the sentence. The dependent variable was the decision time in response to the picture probe.

Sentences were presented in a random order, the same recording being used for the immediate and delayed probe conditions to minimize differences between conditions. All subjects used their dominant hand for the "Yes" response button.

Subjects

Subjects were sixteen volunteers of similar educational background. All subjects were fluent English speakers. None of the men or women who took part in the experiment had participated in previous experiments, and all subjects were naive to the purposes of the experiment.

Stimuli

Thirty "complex" sentences from Experiment 1, plus 6 practice sentences were recorded by a competent reader. Each sentence was preceded by the word "Ready" and the beginning of the sentences were separated by 15s. Two copies of the tape were made. On one copy pulses were recorded on a second channel directly after the noun (initial), on the other copy pulses were recorded on a second channel directly after the final word of the sentence (final).

For each sentence a corresponding slide was produced such that 20 sentences were paired with slides that were totally unrelated to them. The remaining 10 target sentences were paired with two corresponding slides; one was a typical match to an object in the sentence (as determined by Experiment 1), the second was a picture of the same object appropriately modified to fit the sentence context. For example, the target sentence "Helen put it aside because she was in a hurry, even though the *half-eaten fruit* appealed to her" was paired with (i) a picture of an apple and (ii) a picture of a half-eaten apple. All subjects were presented with items from both of these target conditions although each subject was only presented with one picture from a particular pair, i.e. *apple* or *half-eaten apple*. Slides were photographs of simple colour drawings typically taken from children's alphabet books and children's colour dictionaries. (A list of target sentences and pictures is given in Appendix 3).

A test sheet was prepared which listed the 30 target nouns the subject had heard in the sentences. Each noun was preceded with a space in which the subject could write the adjective that had preceded it. The nouns were typed on an A4 sheet in a random order in two columns of 15.

Apparatus

A reel-to-reel tape recorder with in-built pulse mechanism was used. This was connected to a Kodak carousel slide projector via a standard t-scope. The pulse on the tape signalled a ms timer to start and the t-scope to open the projector shutter to display the slide for 1 sec. The carousel then automatically moved the next slide into position. The timer stopped when the subject pressed the "Yes" or "No" response button, and was reset manually by the experimenter.

Procedure

Each subject was tested in one experimental session which lasted approximately 30 min. Subjects were seated behind a table on which the projector was mounted, approximately 3m away from a blank wall onto which the image was projected. The approximate size of the projected image was $1m^2$. Subjects were given the response keys, the bigger "Yes" button they held in their dominant hand, the smaller "No" button in their other (typically left) hand. *Subjects were told that their task was to decide if the picture displayed was a picture of an object in the sentence, regardless of the other information in the sentence or the meaning of the sentence as a whole.* Subjects in the initial condition were also informed that at the point of picture onset they would already have heard all the information necessary to make a correct decision. This was to discourage subjects from waiting until the sentence had been completed before making a response. The task was made clear by example. The subject was then presented with six practice trials with feedback, during which they were encouraged to answer as quickly and as accurately as possible. The subject had an opportunity to ask any questions after the practice before moving on to the experimental session of 30 trials.

After the experimental session was completed subjects were given a surprise recall task. Subjects were presented with the 30 nouns from the experiment and asked to write down as many prenominal adjectives as they could remember. Subjects completed the recall sheets in their own time. As recall scores were often very low, it was stressed to the subjects that the recall test was simply to provide a measure of how much attention they had actually been paying to the sentences as a whole. It was emphasised that it had not been necessary to attend to the whole sentence in order to do the task as they had been asked, and thus a low recall score was not a reflection of poor ability.

RESULTS

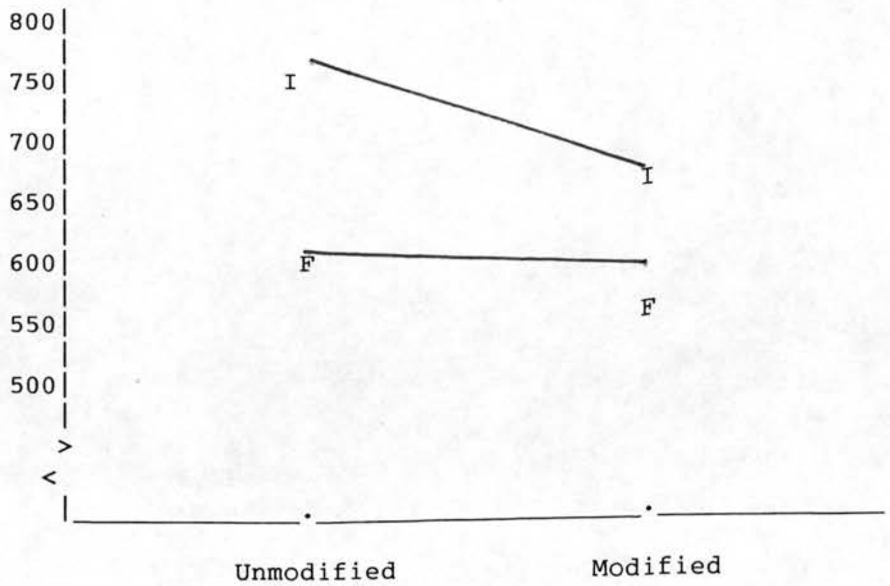
Decision Times

The average error rate was 4.6% (6.9% for positive probes and 3.4% for negative probes). The tendency towards higher error rates for positive probes (also reported by Potter & Faulconer) shows that subjects used strict criteria: being more likely to reject a match mistakenly, than to make one mistakenly. A comparison of the mean error rates presented in Table 1 shows that more errors were made in the initial condition (5.8%) than in the final condition (3.3%). This difference is largely due to more wrong "Yes" responses to filler items in the initial condition, and may be a result of task pressures (for instance the infrequency of "Yes" answers may lead subjects to predict when they are likely to occur). Errors were analysed by fitting error rate to a 2 X 3 probe position (initial and final) X probe type (typical, atypical and filler) logistic model. Thus, proportions of right and wrong answers were compared with a binomial distribution. No main effects or interactions were significant (Chi squared probe position = 3.74, df = 4; Chi squared probe type = 2.65, df = 3; Chi squared interaction = 0.91, df=2). All errors were excluded from further analysis. As "No" responses were included in the experiment only as filler items, so that there would be no incentive to combine the adjective and noun in the task, all "No" responses were also excluded from the analysis. Decision times for "Yes" responses were analysed in a 2 X 2 (probe type X probe position) mixed ANOVA. Average response times and error rates for each condition are presented in Table 1.

Analysis of variance showed that there was a significant main effect of probe position ($F(1,68) = 4.39$, $p < 0.05$), indicating that decision times were significantly faster overall when the probe was presented at the end of the sentence. This may be partly explained by interference between the remainder of the sentence being read while subjects were making a decision in the initial condition. Post-hoc t-tests revealed that subjects responded significantly faster when the unmodified probe was in the final position ($t = 2.088$, $df = 70$, $p < 0.025$), but that there was no difference for modified probes (t = 1.351, df = 70) - see Graph 1. This result is due to the fact that subjects take substantially longer to respond to unmodified probes in the initial condition. The fact that no significant difference was found for modified probes in the final condition may be a result of responses in this condition reaching a floor. This hypothesis is supported by the comparatively low standard deviations found in this condition - see Table 1. The main effect of probe type did not reach significance (F(1,68) = 3.70), this was probably the result of the large standard deviations observed in the data. Neither was there a significant interaction between probe type and probe position (F(1,68) = 1.47).

INITIAL	UNMODIFIED	MODIFIED	FILLER
0.709 SD 0.303 (5.8%)	0.764 SD 0.345 (7.5%)	0.654 SD 0.243 (7.5%)	(5%)
FINAL			
0.590 SD 0.267 (3.3%)	0.605 SD 0.288 (5%)	0.576 SD 0.241 (7.5%)	(1.9%)
	0.686 SD 0.328 (6.2%)	0.616 SD 0.245 (7.5%)	

Table 1
Experiment 2: Average decision times (in ms) of subjects to colour slides which matched (i) only nouns (unmodified), and (ii) nouns and adjectives (modified), in auditory presentation. Decision times are for presentation of the slide immediately after the noun (initial) and at the end of the sentence (final). % proportion of errors per condition are shown in parenthesis.



GRAPH 1
Decision speeds (in ms) for "Yes" responses to unmodified and modified probes in initial and final positions.

These results confirm the main finding of Potter & Faulconer that subjects have little difficulty matching the modified picture to the category name as it appears in context. Subjects are able to make this match immediately after the presentation of the category name.

Materials Analysis

It was not feasible to perform an *ad hoc* materials analysis for Experiment 2 due to the small number of repetitions of each item, the fact that order was not random across items, and the presence of a between subjects factor across relevant comparisons. The average response times for each item in the Initial and Final conditions are presented in Graph 2. The items are presented in pairs, "hammer" then "broken hammer", for ease of comparison. It should be noted that different subjects judged each member of the pair (e.g. "hammer" and "broken hammer"), and different subjects judged the same item in the Initial and Final conditions. Thus each point in the quadruple for "hammer" and "broken hammer" was contributed by a different group of subjects. Thus, any apparent differences between items may simply be due to subject differences.

Despite these difficulties, inspection of the graph reveals a few trends. For all items, decisions in the Final condition were faster than decisions in the Inital condition. In general, decisions in the Final condition follow the same pattern as decisions in the Initial condition, which agrees with the results reported earlier - see Table 1. As can also be seen from Graph 2 there is considerable variation between items with respect to the difference in responses in the Initial and Final condition, and the relative difference being presented with a modified probe made. All these variations may be caused by between subject factors, or order effects, as outlined above. In Graph 2, no individual item stands out as an exceptional case.

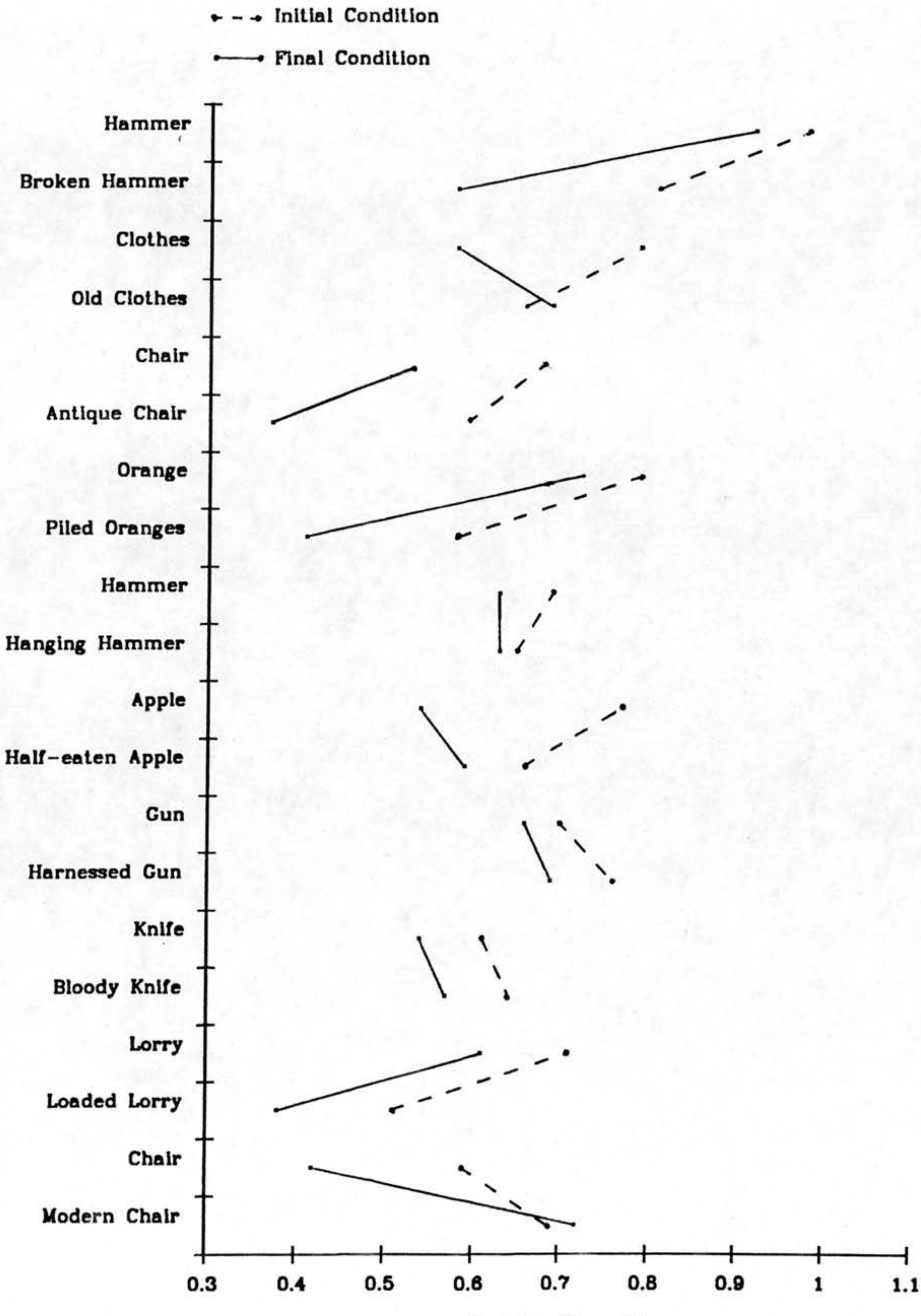
Recall

Subjects were unexpectedly given a list of nouns at the end of the experiment and asked to supply the relevant adjectives. The recall scores were analysed by fitting the proportion of correctly recalled adjectives to a 2 X 3 probe position X probe type logistic model. The analysis revealed that the main effect of probe position was not significant (Chi squared = 0.58, df = 1). There was a significant main effect of probe type (Chi squared = 105.12, df = 2, $p < 0.001$) and the interaction term was also significant (Chi squared = 105.88, df = 3, $p < 0.001$). The significant interaction is a result of recall scores for fillers improving in the final condition, while recall scores for modified and unmodified probes were slightly worse in this condition. The recall scores are summarised in Table 2.

Subjects report, and the results demonstrate, that it is easier to remember adjectives for nouns if you have seen a picture of that noun. This is shown in Table 2 where recall scores for unmodified and modified probes (60% and 72%) were higher than recall scores for filler items (19%) which did not have a matching picture. Recall scores for modified probes were highest, showing that subjects found it easiest to remember the appropriate adjectives when shown a picture

GRAPH 2

Response times for modified and unmodified items in the Initial and Final conditions



which incorporated adjectival qualification. These results are in keeping with the findings of Potter & Faulconer who point out that they can be explained in terms of levels of processing. Thus, the more processing a stimulus is given the more permanent the memory trace laid down by it, and so the easier it is to recall. Contrary to that reported by Potter & Faulconer, in this experiment the position of the probe had no significant effect on recall scores (33% vs 37%).

DISCUSSION

Experiment 2 demonstrated that context is taken into account at a very early stage when making category decisions. There are, however, two possible explanations that must be discounted before it can be asserted that context has as effect on the retrieval of exemplars: the overall matching explanation, and the possibility of swift post-access modification. Each of these alternatives will be discussed in depth below.

Firstly, the results of Experiment 2 can be explained in terms of an overall matching strategy. That is, modified pictures matched the meaning of the sentence twice: one match with the noun, and one with the adjective. So, for example, the picture "half-eaten apple" matches the adjective "half-eaten" as well as the category "fruit". Similarly, the picture "broken hammer" matches the adjective "broken" as well as the category "tool". Thus, the advantages of modified over unmodified probes could be explained by saying subjects responded most quickly to the picture that most matched the sentence, even although subjects were instructed to ignore the sentence context, and Potter & Faulconer found no advantage for modified probes where the adjective occurred in the clause prior to the noun. The fact that there was no significant difference between modified and unmodified probes in the final condition, cannot be taken as clear evidence against the use of such a strategy because of the possibility that responses to modified probes reached a floor in this condition. In order to completely rule out an overall matching strategy, probes would have to be constructed in such a way that the modifier was not visually represented in the probe *independently* from the noun.

	Unmodified	Modified	Fillers	
Initial	62%	75%	16%	33%
Final	58%	70%	23%	37%
	60%	72%	19%	35%

Table 2
Experiment 2: The percentage of matching adjectives recalled when presented with nouns, displayed as a function of probe type and probe position during exposure.

A second possible problem with the interpretation of these results is the fact that context may not have been involved in the *retrieval* of an exemplar, as the exemplar suggested by each condition was the most typical category member. Since there are good reasons to assume that all possible modifications (half-eaten, broken, hanging, harnessed etc.) are not pre-stored, it is possible that the modification required by the context can be quickly achieved after retrieval - quickly enough to give modified instances a quicker response latency in the initial stage. Or, as seems more likely, given the possibility of an overall matching strategy, it is possible that the most typical exemplar (e.g. hammer) was retrieved in both conditions, and that the advantage in the modified condition was solely due to the fact that the picture also provided a good match for the adjective "broken".

Before it can be asserted that context affects the immediate selection of exemplars, it must be demonstrated that modified probes have an advantage over typical probes where the modifier does not match the picture independently from the noun. In addition, the selection of the appropriate category member in the modified condition should be context dependent. That is, the context sentence should suggest an atypical exemplar which would not be routinely accessed in the presence of the category name.

3.3. Experiment 3

To What Do Modified Categories Refer?

Introduction

How does context affect people's understanding of category referents in on-line processing? Are people quicker to recognise an atypical referent of the sentence, or does the category name activate a fixed prototype in on-line processing regardless of context? These are the questions that Experiment 3 is designed to answer. Following closely from Experiment 2, this experiment considers category reference in sentences suggesting atypical referents. The possibility of an overall matching strategy being used in this experiment is ruled out in this task because the adjective is not visually represented in the probe picture independently from the noun (Note 6).

METHOD

Design

The design was identical to that in Experiment 2 save that unmodified instances were replaced with pictures of the category prototype, and modified instances with pictures of the atypical referent. E.g. for the sentence: "He won the *fairground fruit* in a competition" the prototype for FRUIT is "apple" whereas the context modified referent is, the normally atypical instance, "coconut". Information about prototypes and context specific referents was found in Rosch (1975a), Uyeda & Mandler (1980) and Experiment 1.

Subjects

Sixteen new subjects participated in this experiment none of whom had participated in any previous experiments. One subject was excluded from the study as she only gave 4 "Yes" responses; this subject was replaced.

Stimuli

Stimuli were prepared in an identical way to Experiment 2. A new recording was made of 30 sentences and 6 practice sentences. The 10 target sentences all suggested atypical referents and were the same as those used in the experiment reported in Appendix 2. (A list of the target sentences used, and their typical and atypical referents can be found in Appendix 4). Filler sentences and target sentences from Experiment 2 were used as filler sentences in Experiment 3. Thus the filler sentences were of a similar construction to the target sentences, but were not paired with matching pictures.

A test sheet was also prepared as in Experiment 2. The apparatus and procedure were identical to those used previously.

RESULTS

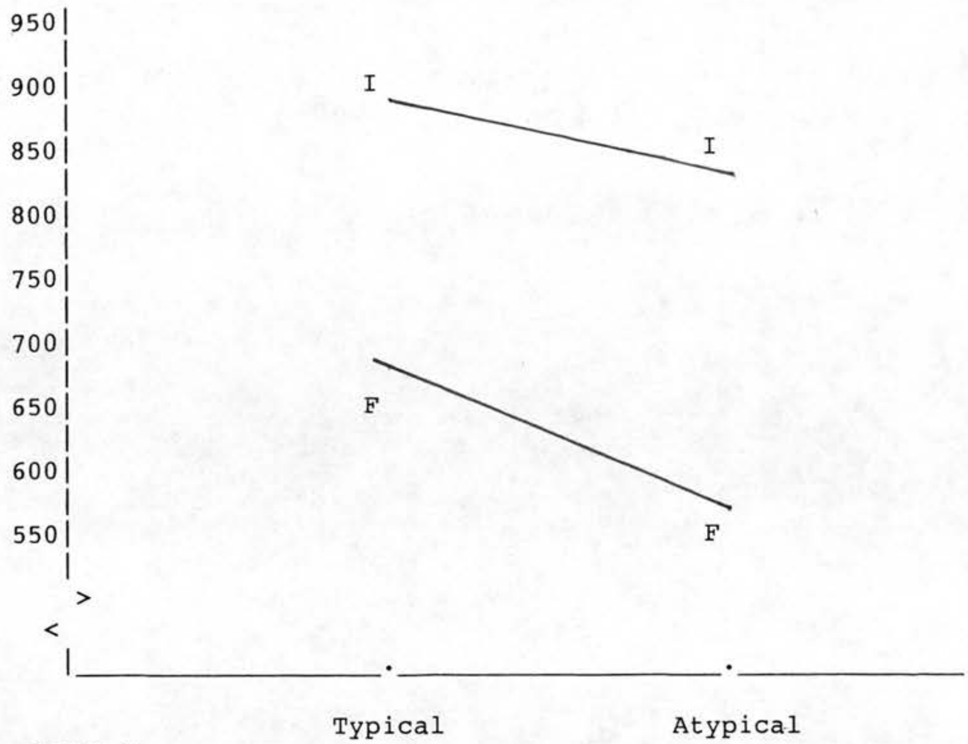
Decision Times

The average error rate was 7.1% (12.5% for positive probes and 4.4% for negative probes). The error rate for positive probes was double that found in Experiment 2. This indicates that subjects found this task more difficult. A 2 X 3 probe position X probe type logistic model was fitted to the data. Analysis revealed that there was no difference between errors in the initial condition (6.2%) and final condition (7.9%) (Chi squared = 1.2, df = 1). There was a significant main effect of probe type (Chi squared = 12.21, df = 2, $p < 0.01$), and a significant interaction (Chi squared = 13.46, df = 3, $p < 0.01$). This shows that subjects made significantly more errors when presented with a typical probe in the initial position. This is consistent with the long decision times for probes in this condition, showing that subjects found it difficult to reconcile a typical exemplar with the atypical context. All errors and "No" responses (filler probes) were excluded from further analysis. Decision times for "Yes" responses were analysed in a 2 X 2 (probe type X probe position) mixed ANOVA. Average response times and error rates for each condition are presented in Table 3.

There was a significant main effect of probe position such that decision times were significantly faster when the probe was presented at the end of the sentence ($F(1,64) = 10.17$, $p < 0.01$). As in Experiment 2 this difference may be partly explicable in terms of an interference effect in the initial condition. Post-hoc t-tests revealed that subjects responded significantly faster to both probe types in the final position (typical: $t = 1.986$, $df = 66$, $p < 0.05$, atypical: $t = 3.138$, $df = 66$, $p < 0.005$), but that subjects responded fastest when the probe was an atypical referent ($t = 1.726$, $df = 34$, $p < 0.05$). As well as high error rates for target items in the initial condition there were very high standard deviations, so although the MS for the main effect of probe position was higher than in Experiment 2, the variance was also much higher. Thus, the main effect of probe type on decision time did not reach significance ($F(1,64) = 1.57$), and there was no significant interaction ($F(1,64) = 0.15$).

INITIAL	TYPICAL	ATYPICAL REF	FILLER
0.862 SD 0.499 (6.2%)	0.891 SD 0.543 (15%)	0.834 SD 0.449 (12.5%)	(2.5%)
FINAL			
0.626 SD 0.246 (7.9%)	0.680 SD 0.287 (10%)	0.571 SD 0.182 (12.5%)	(6.2%)
	0.782 SD 0.444 (12.5%)	0.699 SD 0.364 (12.5%)	

Table 3
Experiment 3: Average decision times (in ms) of subjects to colour slides which (i) matched the typical category referent, and (ii) matched an atypical referent which was context appropriate. Decision times are for presentation of the slide immediately after the noun (initial) and at the end of the sentence (final). % proportion of errors per condition are shown in parenthesis.



GRAPH 3
Decision speeds (in ms) for "Yes" responses to typical and atypical context relevant probes in initial and final positions.

Thus, subjects responded most quickly when presented with a context relevant probe at the end of the sentence - see Graph 3. In the initial condition subjects respond more slowly to both typical and atypical probes than in Experiment 2. This, in conjunction with the high error rates and standard deviations, demonstrates that subjects found this task more difficult. It seems that context relevant exemplars have enough influence to offset the normal advantage of typical instances over atypical instances in categorization decisions. Thus it seems that the subjects took the same amount of time to ignore the conflicting context information for the typical exemplar, as to retrieve the atypical exemplar. When the probe was presented in the final condition context had a stronger effect that typicality. This might suggest that the influence of context increases over time to a greater extent than the influence of typicality.

Materials Analysis

A similar procedure for investigating materials effects to that used in Experiment 2 was adopted here, since the same conditions apply to both.

Inspection of Graph 4 shows that, as in Experiment 2, decisions times are faster for items in the Final condition. One exception to this is for the item "carrots" which is recognised more quickly in the Initial condition. In general, there is a lot of variation in responses and no clear indication that Inital and Final conditions mirror each other, as found in Experiment 2. This may reflect increases in variation when people judge less typical items.

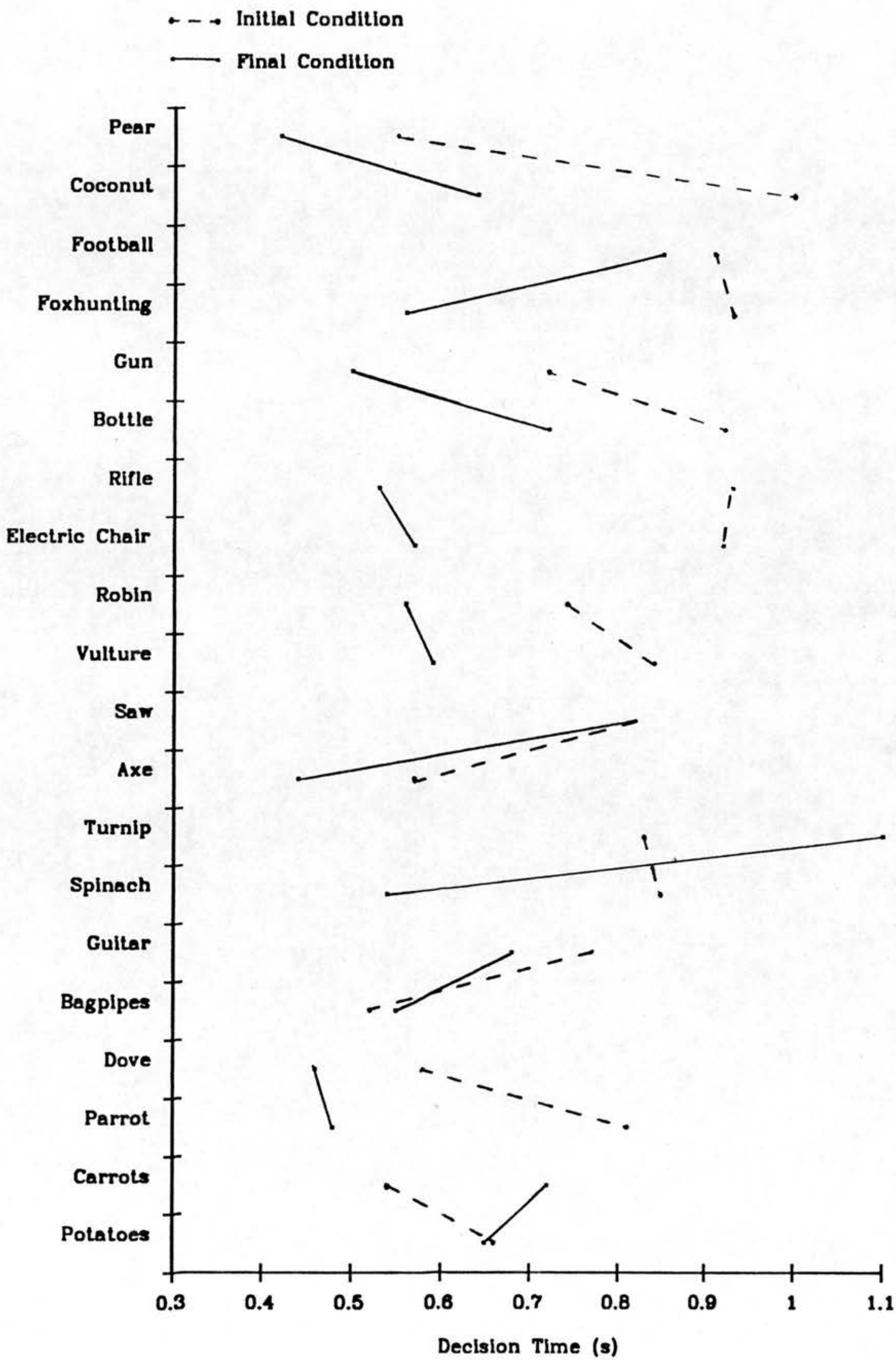
Recall

Subjects were unexpectedly given a list of nouns at the end of the experiment and asked to supply the relevant adjectives. Recall scores were fitted to a 2 X 3 probe position X probe type logistic model. The main effect of probe position was not significant (Chi squared = 0.53, df = 1). There was a significant main effect of probe type (Chi squared = 13.24, df = 2, $p < 0.01$) and a significant probe position X probe type interaction (Chi squared = 13.71, df = 3, $p < 0.01$). The interaction is different from that found for recall scores in Experiment 2, and seems to be reflecting the fact that subjects recall a higher percentage of adjectives for typical probes in the initial condition, and for atypical probes in the final condition. The recall scores are summarised in Table 4.

The results for filler items are the same as those found in Experiment 2. Modified and unmodified probes had a smaller effect on recall scores in this experiment. The higher recall scores in Experiment 2 could be explained by the fact that the category name "tool" was a good recall cue for the exemplar "hammer", since it is a typical exemplar. In turn, "hammer" then served as an excellent recall cue for the probe picture, from which the adjective "broken" could be inferred. Since adjectives were not visually encoded in the picture probes in Experiment 3, it is unlikely that

GRAPH 4

Response times for typical and atypical items in the Initial and Final conditions



subjects *inferred* the adjective in this case. This could account for the lower recall rates for typical (unmodified) instances in Experiment 3. The fact that category names would not provide such good recall cues for atypical (modified) instances could explain why these recall scores were lower. Analysis of the recall scores for Experiment 3 provides some evidence that it is easier to remember adjectives for nouns you have seen a picture of. There is also a tentative suggestion that typical instances are more memorable when presented in the initial condition, whereas atypical instances are more memorable for the final condition. As in Experiment 2, the position of the probe had no effect on recall scores.

	Typical	Atypical	Fillers	
Initial	48%	35%	17%	25%
Final	22%	42%	26%	28%
	35%	39%	21%	26%

Table 4
Experiment 2: The percentage of matching adjectives recalled when presented with nouns, as a function of probe type and probe position during exposure.There were no probe matches for filler items.

Discussion and Reanalysis

In Experiment 2 the only significant differences between the modified and unmodified probes was when subjects responded significantly faster to the modified probe in the initial condition. In this experiment, the only significant difference between probes occurred when subjects responded fastest to the modified probe in the final condition. In both experiments the fastest response times are for the modified probe in the final condition, and the slowest response times are for unmodified probes in the initial condition. One possible explanation for the advantage of modified probes in the final condition in Experiment 3 is the fact that more context information was present in the target sentences *after* the occurrence of the target noun in this case. This was due to the constraints of sentence construction, in conjunction with the fact that it is more difficult to make a sentence suggestive of an atypical exemplar than a typical one. Hence, more context information had to be included in the target sentences of Experiment 3. It was assumed that this additional context information would not change which exemplar best fit the category in the context, as the extra information was always consistent with the adjective-noun pair. Barsalou's idea of complex concept formation in all contexts, consonant with the results of Experiment 1, lends some support to this assumption (Note 7).

INITIAL	TYPICAL	ATYPICAL REF
0.914 SD 0.585	0.937 SD 0.654	0.891 SD 0.507
FINAL		
0.606 SD 0.224	0.613 SD O.241	0.598 SD 0.205
	0.771 SD 0.516	0.741 SD 0.410

Table 5
Experiment 3: Average decision times (in ms) of subjects to colour slides which (i) matched the typical category referent, and (ii) matched an atypical referent which was context appropriate. Decision times are for presentation of the slide immediately after the noun (initial) and at the end of the sentence (final).

However, in order to assess the possible biasing effects of context information included after the noun, a small normative study was done. Fifteen new subjects were given a list of the 10 partially completed target sentences. Like Experiment 1, they were asked to provide a best fit, and second best fit category instance for the partially completed sentences. The formation of these norms provided the context appropriate exemplar from the information available to the subject at the occurrence of an initial probe. It was found that 4 sentences may have suggested different referents (see Appendix 4) at the point of the occurrence of the noun, while all other sentences suggested the same referent at this point as at the end of the sentence. The responses for those sentences suggesting different referents were excluded, and the data reanalysed as before. There was no difference between these ANOVA results and the ANOVA results for the whole data set: only the main effect of probe position was significant ($F(1,38) = 6.36$, $p < 0.025$). The only difference between the analyses is that post-hoc t-tests showed no significant difference between typical and atypical probes in the final condition ($t = 0.211$, $df = 20$). This result suggests that such a difference may have been solely due to the integration of context material in the final condition, which appeared at the end of the sentence. The means and standard deviations for sentences suggesting the same exemplars at the initial and final probes are presented in Table 5. Thus, it can be concluded that in Experiment 3, subjects responded equally quickly to typical and atypical probes, but much faster to probes presented at the end of the sentence. Like the results reported for Experiment 2 the longest response latencies were for the unmodified (typical) probe in the initial condition.

DISCUSSION

What has clearly emerged from these experiments is that context is automatically taken into account even in tasks where subjects are instructed not to do so. Experiment 3 has shown that the influence of context cannot be explained in terms of an overall matching strategy. It has also demonstrated that context is involved in the retrieval of category exemplars, as this task required the integration of context information as a prerequisite for making the correct match quickly.

The surprising fact that has emerged from the experiments is that relatedness appears to have no effect on categorization decisions, when the qualifying information is not represented independently (as in Experiment 3). The atypical, context-relevant exemplar, is matched as quickly as the category prototype, even when the probe is presented immediately after the target word. The data from experiments 2 & 3 seems best explained by arguing that current contextual appropriateness is at least as important as previously learned appropriateness which typicality represents. (This notion could be explored further by comparing how subjects respond to atypical items that are contextually inappropriate, and typical items that are appropriate in context). The clear result that emerges from these experiments is that context is integrated immediately a word is read, and is involved in selecting a category referent.

3.4. Context and Theories

Experiment 3 clearly demonstrated that context influences the understanding of a category term, and that that influence cannot be ascribed to a post-access period. This conclusion makes it unlikely that coherence is a result of conceptual representation *per se*. The alternative view, which is consistent with immediate contextual modification, states that coherence is a result of the relations between concepts that are computed relative to context.

Medin & Watenmaker (1987) argue that people not only notice the correlational structure of the world (that fins and gills go together), but they also notice the reasons behind such correlations (survival in an aquatic environment). This kind of processing can also be used to explain social categorization phenomena like illusory correlation. Hamilton & Gifford (1976) found that people wrongly infer that two unusual events or stimuli which co-occur are causally related. Moreover, people maintain beliefs supporting that connection over time.

Explaining coherence in terms of theories does not completely solve the problem of context effects. This is because the selection of a theory in a particular situation will be largely context dependent. As coherence emerges from the links between items, it cannot be said to exist independently of the theory under consideration. It does not follow from this that coherence no longer places any constraints on possible concepts. On the contrary, coherence may well provide a test of which theory generates the most likely meaning for a term in a particular context. As theories are much more general than concepts they should not change much from context to context, once they are fully developed. It is the theory which is chosen that should vary with the demands of context. The fact that theories do not undergo considerable change with context gives a sense of stability to word usage.

I have argued that people ascribe membership to a category as a function of the perceived relationship of that category to the current context. If this is accepted, it can be seen how Barsalou's notion, that complex concepts are routinely formed when processing sentences, might hold. Firstly, when making membership judgements about complex concepts like "pet fish" the fact that the two concepts PET and FISH are in a relationship is made explicit. In a context sentence, for example: "His fish was the best pet he had ever had", the relationship between PET and FISH is implicit. In this second case it is our theory about the possible meaning of the sentence that links PET and FISH together. In both these cases the relationship is understood in terms of the same theory. Or, in other words, a complex concept (explicitly acknowledging the relationship) is formed in both cases. It may at first sight seem ridiculous to talk about linking concepts like PET and FISH in a "theory" since they seem so obviously related. What is meant by a theory is, however, the implicit assumptions and beliefs that have been collected over time and built into a framework for understanding the world. If people did not have such assumptions and beliefs about fish and pets they would simply fail to understand what the sentence meant.

The question of how relevant attributes are selected in different contexts can now be construed as the result of forming different complex concepts, or holding different relational theories, in each case. An understanding of how complex concepts are formed, that is, why certain attributes are relevant in that context, now seems crucial for the development of a theory of concepts.

CHAPTER 4

OVEREXTENSION OF CONJUNCTIVE CONCEPTS

4.1. Conceptual Combination [1]

If, as argued previously, the solution to contextual change must be intensional then looking at noun-noun combination may provide clues to how and why these changes occur. This is because complex concepts can be conceived of as a special kind of context effect where each word acts as context for the other.

Those who hold to a Classical View have attempted to explain concept combination in terms of combinations of definitions. It has been argued from the fact that language is compositional that semantic representations must also combine compositionally (Dowty, Wall & Peters, 1981). Semanticists have further argued that this provides the only adequate account of the productivity of language and thought. The Classical View of concepts is that the formal semantic account is an appropriate model of how people actually make categorization decisions. In the specific case of conjunctive concepts, the Classical theorist proposes the "Boolean" hypothesis. That is, that the members of the constituents are set-theoretically conjoined to form the members of the conjunctive concepts. Membership of the complex concept is a Boolean function of membership of the constituent concepts. The Boolean hypothesis predicts that, for example, an item is judged to be a SPORT WHICH IS ALSO A GAME if and only if it is judged to be both a SPORT and a GAME.

In a Classical account, there is no room for graded membership: category membership is construed as being all or none. Classical theorists thus claim that the processes underlying typicality effects are separate from those underlying membership judgements. For instance, Osherson & Smith (1982) have concluded that "the theoretical machinery needed to explain typicality findings is different in kind from that needed to explain compositionality" (p. 317). This has been called the Binary view (Hampton, 1988). If the Binary view holds, then context no longer

[1]The work reported in this chapter was carried out in collaboration with Nicholas Chater, Centre for Cognitive Science, University of Edinburgh.

presents a major theoretical challenge. The underlying representational structure can remain intact, and all that would be needed would be the inclusion of a subsidiary process to adjust decisions to fit a particular context. If the Binary hypothesis is false then the fact that typicality judgements change with context has serious representational implications.

The second theory of conceptual combination commonly proposed is that of Fuzzy Set Theory. This theory of combination is more acceptable to prototype theorists as it allows that membership can be graded. In response to criticisms from Osherson & Smith (1981) and others, Zadeh has clarified his conception of a prototype in Fuzzy Set Theory: "A prototype is not a single object or even a group of objects in A. Rather, it is a fuzzy schema for generating a set of objects which is roughly coextensive with A" (p.293, Zadeh, 1982). Thus, Zadeh is emphasizing the need to consider conceptual combination as *intensional* change.

James Hampton has argued against the Boolean position. In 1988 he reported findings which suggest that subjects do not judge membership in conjuncts according to a Boolean hypothesis. Rather, subjects include items as members of a conjunct category they have previously excluded from one or both constituent categories. On the basis of these findings Hampton proposes a Unitary view that membership and typicality are "attributed to a single underlying factor" (p 12). I shall now consider Hampton's experimental paradigm in some detail.

In Hampton's Experiment 4 subjects assessed the membership and relatedness of lists of items to categories (stage 1). A week later subjects assessed the membership and relatedness of the same items to a conjunction of two of the categories (stage 2). For example, at stage 1, a subject might decide that "skateboard" was a member of the category VEHICLE (+), but not of the category MACHINE (−). At stage 2, the subject might decide that "skateboard" is not a VEHICLE WHICH IS ALSO A MACHINE (−). This triple of responses is represented (+−−).

Hampton instructed subjects to rate a number of items for typicality and membership (using a seven point rating scale from -3 to +3) according to the following instructions:

> For each example first decide whether you would answer 'Yes' or 'No' [membership], and then select one of the corresponding positive or negative values to indicate the strength of your choice [typicality of members and relatedness of non-members]. If you are unable to decide, use the value zero, but avoid using this as much as possible.

comments in [ ] mine

p. 16

The design of the experiment and the instructions pose problems for the interpretation of Hampton's data. It is a serious criticism of this experiment that the Unitary hypothesis (that typicality and membership are attributable to the same underlying factor) is inadvertantly built in to the rating scale used in the experiments. Hence it can be argued that Hampton's experiments assume the Unitary hypothesis rather than confirm it. Further, it is unclear that these experiments undermine the Boolean hypothesis because there are other possible interpretations of Hampton's results. I will illustrate these alternative interpretations below by examining the rating scale, design and the analysis of Hampton's original experiments.

Rating Scale

The rating scale used may bias subjects' responses towards those predicted by the Unitary hypothesis for the following reasons:

1. In order to assess whether or not typicality and membership are aspects of the same phenomenon, it is necessary to have independent measures of both. Since the rating scale measures typicality and membership together it might be expected that the membership task will distort the typicality ratings, and the typicality task

will distort the membership ratings in the following ways:

(i) The requirement to judge typicality may bias subjects to adopt a similarity based categorization strategy. Hence, the results might be biased towards those predicted by the Unitary hypothesis, as Hampton has since pointed out (personal communication, 1988)

(ii) The scale does not allow items to be rated in a way which violates the Unitary hypothesis. Thus it is impossible to give a non-member a typicality/relatedness rating higher than -1. Similarly, it is impossible to give a member a typicality/relatedness rating lower than 1. Hence members are necessarily given higher typicality/relatedness ratings than non-members. A violation of the Unitary hypothesis would be obtained if some non-member was rated as more typical/related than some member. For example, subjects might want to rate "tomato" as a typical VEGETABLE, although the subject knows it is a FRUIT. In particular, they might want to give "tomato" a higher typicality/relatedness rating than an atypical vegetable such as "yam". The subject cannot simultaneously respect intuitions about typicality and membership. If membership takes precedence, then the subject is forced to give a lower typicality/relatedness rating to "tomato" (say, -1 for related non-member) than for "yam" (say, +1 for atypical member). Alternatively, if typicality takes precedence, then "tomato" must be given a higher typicality/relatedness rating than "yam". This means that if "yam" is judged to be a vegetable, then "tomato" will be judged to be a more typical vegetable. On the other hand, if "tomato" is judged to be related non-member of VEGETABLE, then "yam" must also be a (less related) non-member. A Unitary theorist might respond that there are no such cases. However, this just *is* the Unitary hypothesis.

2. Whereas the previous point relates to judging membership of constituent categories, a different problem emerges in judging membership of conjuncts. In this case, subjects might feel unhappy about giving a negative rating if the item was a good member of one of the conjunct categories. For example, "chess" might

be judged not to be a member of SPORT. However, if the subject considered "chess" a very good member of GAME it might be included by the subject in the category SPORT WHICH IS ALSO A GAME. Subjects may overextend their categories because they wish to express that an item which is a member of one constituent category is a better member of the conjunct than an item which is not a member of either constituent category (such as "watching television"). To guard against this possible source of spurious non-Boolean effects a wider range of response options could be given to subjects at Stage 2. For example, subjects could be given the option of rating "chess" as: A GAME WHICH IS ALSO A SPORT, JUST A GAME, JUST A SPORT, or NEITHER A GAME NOR A SPORT. This gives subjects the opportunity to distinguish between items they consider to be members of one category, and items that are not members of either category. Thus, at stage 2, "chess" can be judged to be JUST A GAME.

Design and Analysis

In Hampton's experiment the number of responses expected in each response class (e.g. ++-, -++) was estimated. As Hampton points out, certain non-Boolean responses would be expected even if subjects use a Boolean combination rule, given i) that subjects forget their previous responses with probability $u < 1$, and ii) each membership judgement for the constituents may be probabilistic. For example, a subject might judge an item to be a SPORT and a GAME at Stage 1. At Stage 2, the subject might not remember the previous judgements, and now decide that the item is not in fact a GAME. If the subject is using a Boolean strategy, the item will not be judged a member of the conjunct. Hence, non-Boolean results do not necessarily imply non-Boolean combination.

Since there is no direct measure of how much subjects remember, Hampton has provided an estimate of a remembering parameter u. This is the probability of subjects remembering their previous judgements. Hampton explains:

The parameter u was estimated from the two observed frequencies for +++ and ++−, and constrained to lie between 0 and 1. If it could not be estimated it was set to zero.

p 17

Using his estimate for u, the probability of each fresh membership judgement, was derived. For example, Hampton notes that the probability of classifying an item as a member of the conjunct, when it had previously been classified as a member of the first constituent but not of the second (+−+) is:

$$p(+-+) = s_1(1-g_2).[u^2.(0) + u.(1-u).g_1 + (1-u).u.(0) + (1-u)^2.s_1.g_1]$$

p. 17

s_1 = probability of an item being rated as a sport when rated first.

g_1 = probability of an item being rated as a game when rated first.

g_2 = probability of an item being rated as a game when rated second.

u = probability of remembering your classification at stage 1.

Using formulae of this sort Hampton was able to estimate the number of responses that should be obtained in each response category, assuming subjects use a Boolean conjuction strategy. This allowed a comparison of the results obtained with estimates on the basis of the Boolean model, allowing an assessment of the degree to which subjects actually used a Boolean strategy.

The estimate of u was based on the observed frequencies of (+++) and (++−) responses alone. Thus, only a small fraction of the data was used in the estimation, so it is unclear how reliable the estimate is. A further complication is that Hampton claims that u "is also estimating the effects of individual differences in concept definitions and any other factors which may reduce the independence of the subject's ratings at each stage" (p 17). It is unclear how u could meaningfully fulfil these functions. These difficulties can be circumvented by introducing an appropriate control condition, to estimate directly subjects' performance given a Boolean strategy (see Experiment 6).

If the variable of rating order of s and g (which Hampton found to have no significant effect) is omitted from the formulae, then it is found that the probability of an underextension, p(++−), is equal to the probability of an overextension, p(−−+) + p(−++) + p(+−+), independent of the particular values of s, g and u (Note 8). Hence, if the number of overextensions and underextensions is significantly different the Boolean hypothesis is disconfirmed. Of course, if the number of overextensions and underextensions are not significantly different, this does not allow the differentiation of the Unitary and Binary hypotheses, i.e. whether typicality and membership decisions are independent or have the same underlying basis. However, in as much as the overextensions and underextensions differ this provides a simple measure of the degree of non-Booleanness. In the experiments reported below, the significant differences between overextensions and underextensions provide the basis for the argument that concept combination is non-Boolean.

In Hampton's Experiment 4, a linear regression model, which predicts the rating of a conjunction from the ratings of its constituents, is given. The validity of linear regression depends on the data being approximately normally distributed along each dimension. Since Hampton's data approximate to a 'U' shaped distribution (the extreme values, −3 and +3, were the most popular responses and subjects were discouraged from using the centre value 0), the large R values could be due to the fact that the data points are clustered at each end of the regression line.

These points imply that further experimental work is required to firmly establish Hampton's findings. The following set of experiments is designed to fulfil this role. Experiment 4 is a replication of Hampton's Experiment 4. This is compared with three further experiments. In Experiment 5, subjects make only typicality judgements for items. Judgements are made on a seven point scale where no indications are given to where the membership boundary should lie. Hence, subjects are given a free hand in judging typicality of items independently of membership constraints in the task. In Experiment 6, subjects make membership judgements only. This should obviate the confounding of membership and typicality judgements. In Experiment 7, a wider range of response options is given, to allow subjects to judge an item to be a member of just one category of a conjunction.

4.2. Experiment 4

Hampton Reconsidered

Experiment 4 is a replication of the principle features of Hampton's (1988) Experiment 4. The same task was used, and the rating scale and items were identical to Hampton 1988, Experiment 4. Although Hampton reported an effect of order in the category name (A BUILDING AND A DWELLING versus A DWELLING AND A BUILDING) in this experiment the order of constituents in each conjunct was counterbalanced. This was because the concern of this experiment is to assess the magnitude of overextensions and underextensions, not to replicate Hampton's work on the head-modifier relation. The order of rating of categories at stage 1 and stage 2 was random. The analysis of the data is based on the relative number of overextensions and underextensions, as argued above. The purpose of this partial replication is to provide data consonant with Hampton's and appropriate for comparison with the other experiments in this series.

METHOD

Design

The experiment is a within subjects design. The within subjects factor is response type (underextensions and overextensions). The same subjects rated items at stage 1 and stage 2. Items were treated as a random factor. The same random order of items within a list was used for all subjects at both stages. The order in which the lists were rated was fully randomized across subjects. The role of categories as head and qualifier was counterbalanced.

Subjects

10 subjects participated in Experiment 4. All subjects were in full time education

in Edinburgh and took part in the experiment as a course requirement. Two subjects were excluded from the experiment due to absence at one of the sessions. The sexes were approximately equally represented in the group and subjects were naive to the purposes of the experiment.

Materials

Materials were identical to those used by Hampton (1988). Six conjunctions were used: Machines - Vehicles, Furniture - Household Appliances, Pets - Birds, Buildings - Dwellings, Food - Plants (or part of a plant) and Weapons - Tools. For each conjunction a list of 16 items was available. This list contained items belonging to both conjuncts, neither conjunct, or only one conjunct. Items were typed in random order on a sheet headed by the appropriate category name e.g., Machine, Vehicle, Machine which is also a Vehicle, Vehicle which is also a Machine. To the right of each item was a seven point scale from −3 to +3. The instruction sheet included an appropriate completed example not used in the experiment.

Procedure

The experiment was conducted in two stages, a week apart. In the first stage, subjects were given a booklet containing 12 pages of test items, presented in a different random order for each subject. Each page was headed by one of the constituent categories e.g., Bird. Subjects were asked to decide the membership and typicality of the items. Firstly, subjects had to decide if the item was a member of the category. If the item was judged to be a member, it was given a '+' rating. If the item was judged not to be a member, it was given a '−' rating. If undecided the subject could use '0', but this was discouraged. Secondly, subjects had to decide how typical members were of the category, by choosing +1, +2 or +3, where +3 is a very typical member. If the item was a non-member, subjects indicated how related it was to the category, by choosing −1, −2 or −3, where −1 is the most related non-member. Subjects completed the ratings at their own pace. Stage 2 was the same as Stage 1 save that subjects rated the 6 conjunctions rather than the 12 constituent categories. The ordering of the conjunctions ("Pets which are also Birds" versus "Birds which are also Pets") was balanced across subjects such that half the subjects received one random ordering and half the reverse ordering.

RESULTS AND DISCUSSION

As argued above, a significant difference between the number of over and underextensions a subject makes is indicative of a non-Boolean combination strategy. Summed over all subjects 2.9% of the responses were underextensions and 15.8% were overextensions. That is, subjects were more likely to classify an item as a member of a conjunct when not a member of a constituent, than to classify an item as a non-member of a conjunct when a member of both constituents. A within subjects t-test found this difference to be significant at $p < 0.05$ ($t = 5.93$ (5)). This clearly replicates Hampton's finding that subjects exhibit non-Boolean responses to membership decisions in conjuncts. Subjects preferred to overextend rather than underextend their concepts. The pattern of overextensions and underextensions is summarized in Table 6. Each entry in the table indicates the mean response to the conjunction for the corresponding values of the head and qualifier. For example, all the responses in which the head was rated −2 and the qualifier was rated +3 at stage 1 are considered together. The mean value of the corresponding conjunctions at stage 2, in these cases, was found to be 0.47.

A strictly Boolean model predicts that only items judged to be members of both constituent categories (the entries in the top left quadrant of Table 6) should be judged members of the conjunction (i.e. be given a positive rating). If subjects showed a tendency to underextend, then items that were judged members of both constituents would be given "−" ratings for the conjunct. As can be seen from Table 6, such a tendency was not evident. On the other hand, if subjects' tendency was to overextend, then positive values would be found for conjunctions where one or both constituents were judged to be non-members. There are several examples of overextensions in Table 6 where the head noun was given a rating of -1 or -2, but the qualifier noun was judged to be a member.

Table 6

Mean ratings of the conjunction for each value of the head constituent and the qualifier constituent, collapsed across categories, for Experiment 4. Only values for n > 4 are shown.

						QUALIFIER	
	+3	+2	+1	0	-1	-2	-3
+3	2.77	2.08	1.62	0.83	-0.03	-0.79	-1.9
+2	1.94	1.11	1.64		-0.8	-0.29	-1.5
+1	1.87	0.6	0		-0.6		-1.45
0							
-1	0.2	0.54	0.11		-1		-2.71
-2	0.47	-1.57					-2.11
-3	-1.34		-1.18	-1.33	-1.78	-1.71	-2.82
HEAD							

There is a marked assymetry between the influence of the categories depending on whether they are in a head or qualifier position, similar to that reported in Hampton (1988). Positive ratings for conjunctions are found when the head noun had a rating of -1 or -2, and the qualifier noun had a positive rating. This indicates that the qualifier is more important in determining conjunct membership than the head. Thus, if an item is not a member of the qualifier category, it is unlikely to attain membership of the conjunct. This effect is independent of the particular categories rated, since all categories were presented in both qualifier and head positions.

Hampton notes that the Guppy effect (a Guppy is a better PET FISH than it is a PET or a FISH) is hard to find. That is, membership rating in the conjunction is seldom greater than in either constituent. There were 33 instances of the Guppy effect (4.3%) in Experiment 4, similarly a low figure.

As argued above, it is as yet unclear whether the source of these non-Boolean effects is the use of a non-Boolean combination strategy. Experiments 5 and 6 are designed to ascertain the possible contributions of typicality judgements and membership judgements to non-Boolean effects.

4.3. Experiment 5

Typicality Ratings of Conjuncts

Introduction

Experiment 4 provided a replication of Hampton's results, demonstrating that overextension is a robust effect. The purpose of Experiment 5 is to examine whether typicality judgements of an item in two constituent categories predicts the typicality judgement for the item when the categories combine. The results from this experiment are not directly comparable with those of Experiment 4, where a clear category boundary is defined on the scale. Experiment 5 will give some idea of the difference between this task, which has no membership constraints, and the one used formerly. It will also provide a measure of how related typicality judgements for constituents are to judgements for conjuncts.

METHOD

Design

The experiment is of a similar design to Experiment 4. The within subject factor is typicality rating (1 - 7). The stimulus materials are identical in all other respects to those used in Experiment 4.

Subjects

9 subjects participated in the experiment as part of a course requirement. All subjects were naive to the purposes of the experiment, and had never given typicality judgements previously.

Procedure

The procedure was the same as Experiment 4, only the instructions to subjects were different. Subjects were instructed as follows:

> ... For each example decide how well you think it fits the category. Ring 7 if you think the item fits the category very well, and ring 1 if you think it doesn't fit at all, using the intermediate values as appropriate. If you are unfamiliar with any of the examples cross them out.

Subjects were then given two examples from a category not used in the experiment. These examples illustrated classification as 7 and 1 only. Subjects completed the judgements in their own time.

RESULTS

Ratings

A simple method which set the category boundary to the mid-point on the scale was adopted. This method rendered the decisions 7, 6 and 5 as denoting membership, and 1, 2 and 3 as denoting non-membership. The typicality rating 4 was said to denote uncertainty of membership. When this method was used it was possible to calculate the percentage of overextensions and underextensions as 14.6% and 1.2%, respectively. This figure is not meant to reflect the fact that subjects were "overextending". It simply demonstrates that placing a simple criterion on typicality judgements can produce results very similar to those found when subjects judge membership and typicality (as reported in Experiment 4.

Table 7

Mean typicality ratings of the conjunction for each value of the head constituent and the qualifier constituent for Experiment 5. Only values for n > 4 are shown.

	QUALIFIER						
	7	6	5	4	3	2	1
7	6.86	6.29	5.46	5.64	4.85	3.08	3.67
6	6.4					3.33	3
5	6.15			4.86		4.13	2.09
4	6.2	5.75	5.8				1.75
3	4.25				3.8		2.4
2	3.44	4	2.38	2.67	3.5		2.1
1	2.64	2.37	2.85	3	1.43	2.17	1.07
HEAD							

The number of triples which exhibited the guppy effect (that is typicality in the conjunction was rated higher than typicality in both constituents) was 61 (7.1%). That is considerably higher than was found in Experiment 4 (4.3%). A post-hoc Mann-Whitney U test indicated that this difference was not significant at the 5% level ($U(8,9) = 19.5$). Table 7 shows mean typicality ratings for conjunctions for each typicality value of the head and qualifier.

In order to compare the results obtained with the pattern of results found in Experiment 4, the typicality ratings were expressed as membership ratings on a scale of +3 to −3. Thus the value for head and qualifier ratings of 7,7 = 6.86, is expressed as a head qualifier rating 3,3 = 2.86. These expressions of the typicality ratings can be found in Table 8.

It is interesting that, in a task where subjects are not asked to make membership judgements, and where no category boundary was set, simply ascribing the mid-point as the category boundary produces results so similar to those found in

Experiment 4. This supports the notion that membership and typicality may indeed be closely linked. Or that typicality ratings strongly bias membership judgements. One difference which does emerge is that there is no head-qualifier effect (see Table 8). Instead, positive judgements are overextended equally for head nouns and qualifier nouns. The pattern of missing data which does not concentrate around the centre value as it has done previously, may indicate a greater variation between subjects in their criteria for category membership. It is possible that between subject differences in the use of the scale may also be obscuring any head-qualifier effect.

Table 8

Mean typicality ratings of the conjunction for each value of the head constituent and the qualifier constituent for Experiment 5, expressed on a scale where 7 corresponds to +3, and 1 corresponds to -3. Only values for n > 4 are shown.

HEAD	QUALIFIER +3	+2	+1	0	-1	-2	-3
+3	2.86	2.29	1.46	1.64	0.85	-0.2	-2.33
+2	2.4					-0.67	-1
+1	2.15			0.86		0.13	-1.91
0	2.2	1.75	1.8				-2.25
-3	0.25				-0.2		-1.6
-2	-0.56	0	-1.62	-1.33	-0.5		-1.9
-1	-1.36	-1.63	-1.15	-1	-2.57	-1.83	-2.93

In conclusion, judgements for typicality are reasonable predictors of subjects' judgements on a task incorporating membership and typicality. Typicality ratings for constituents are fairly good predictors or typicality ratings in conjuncts. Biasing effects extending positive values may have been exaggerated in the above data due to subjects placing category boundaries at different typicality levels.

4.4. Experiment 6

Are Membership Judgements Non-Boolean?

Introduction

In this experiment subjects are required to make membership judgements only. This should eliminate any non-Boolean effects due to the confounding of typicality and membership judgements. In addition, a control group is introduced at stage 2. This allows a measure of how many overextensions would result if subjects were using a purely Boolean combination strategy. Hence, the control group provides a baseline measure of overextension and underextension with which the experimental group can be compared. The control condition provides a direct estimate of non-Boolean effects that are not a result of conceptual combination, thus avoiding the problems of estimating u. It also provides an experimental check on the formula which states that if subjects use a Boolean strategy there will be no difference between the subjects' tendency to overextend and underextend.

METHOD

Design

The experiment is a 2 X 2 mixed design. The within subject factor is response type (number of overextensions and number of underextensions). The between subject factor is group (experimental group or control group). In all other respects the design is the same as Experiment 4.

Subjects

18 subjects took part in the experiment. All subjects were full time students in Edinburgh. One subject from the control group was absent at stage 2 and thus excluded from the experiment. In order to keep the design balanced, a subject was chosen at random from the experimental group, and excluded from the study. The

sexes were approximately equally represented, and subjects were naive to the purpose of the experiment.

Procedure

The procedure was identical to Experiment 4, except that subjects were only asked to make membership judgements, by circling "Yes" or "No" beside each item. If subjects could not decide whether an item was a member of the category they were instructed to leave the item blank. However, as in Experiment 4, this option was discouraged. If a subjects did not understand a word, they were instructed to cross it out. The control subjects completed stage 1 in the same way as the experimental group, at stage 2 the control group simply repeated stage 1.

The control group's responses at stage 2 were Booleanly conjoined, after the experiment, to yield response triples comparable to those in the experimental group. Suppose that a subject rated an item as a member of one constituent categories at stage 1, but not the other (+−), then rated the same item as a member of both categories at stage 2 (++). A Boolean model of conjunction would imply that the subject would rate the item as a member of the conjunct at stage 2 (+). The response triple (constituent 1, constituent 2, conjunction) can now be constructed. In this case it is (+−+), which is an overextension. For example, at stage 1 a subject might judge "snooker" to be a SPORT, and also a GAME (++). At stage 2, the same subject might now judge "snooker" not to be a SPORT (−), but still to be a GAME (+). Boolean conjunction of the subject's responses at stage 2 yields (−). Hence, the subject is deemed to have judged "snooker" to be a SPORT, a GAME, but not a SPORT WHICH IS ALSO A GAME (++−). This is an underextension.

This procedure provided an estimate of the number of overextensions and underextensions that might be expected to result from subjects forgetting their previous responses and changing their minds. This allowed a comparison of the number of overextensions and underextensions in the experimental and control groups. Significantly more overextensions in the experimental group would imply that subjects were not using a Boolean method of conceptual combination.

RESULTS AND DISCUSSION

A 2 X 2 (group X under/over extension) mixed ANOVA was performed. Percentage overextensions and underextensions for each group are presented in Table 9. There was a significant main effect of group ($F(1,14) = 5.49$, $p < 0.05$). The main effect of response type was also significant ($F(1,14) = 14.3$, $p < 0.01$), demonstrating that subjects made significantly more overextensions than can be accounted for by factors such as forgetting previous responses and changing one's mind. There was a significant interaction term ($F(1,14) = 12.96$, $p < 0.01$) as a result of the experimental group making significantly more overextensions and significantly less underextensions than the control group. The number of overextensions and underextensions per category are shown in Table 10.

Table 9

Overextensions and underextensions for a control and experimental group in Experiment 6. Overextensions and Underextensions are expressed as a percentage of the whole.

	Overextensions	Underextensions	
Experimental Group	12.0%	3.1%	7.6%
Control Group	5.3%	4.2%	4.8%
	8.7%	3.6%	

As expected there was no difference between the number of overextensions and underextensions in the control group (overextensions 5.3%, underextensions 4.2%, post hoc t-test, $t = 0.63$, $df = 5$). The number of overextensions in the experimental group was less than in Experiment 4 (15.8% vs 12%) which suggests that typicality decisions may have influenced subjects' judgements in Experiment 4. If the percentage of overextensions made in the control group is taken as a baseline, the suggestion is that subjects used a non-Boolean decision strategy for around 6% - 7% of responses. As discussed in section 4:1 a remaining source of task specific "overextensions" may be the lack of response options at stage 2.

Table 10

Number of overextensions and underextensions per category for Experiment 6.

Category	Overextensions	Underextensions
Furniture and Household Appliances	31	3
Food and Plants	18	4
Weapons and Tools	19	6
Buildings and Dwellings	3	6
Machines and Vehicles	11	4
Birds and Pets	10	1
	--	--
	92	24

4.5. Experiment 7

Choice Factors in Non-Boolean Categorization

Experiment 7 is designed to investigate possible effects on subjects' tendency to overextend membership decisions for conjuncts. It has been argued previously that subjects may make overextensions simply because they wish to express that an item which is a member of one of the constituent categories is somehow more related to the conjunct than an item which is not a member of either constituent category, nor of the conjunction. In order to test for this possible source of spurious non-Boolean effects a wider range of response options was introduced at stage 2 of Experiment 7. Experiment 7 is a variant of Experiment 4, in this case subjects were asked to make membership judgements, and were given four response options at stage 2.

METHOD

The method and procedure were the same as those used in Experiment 4, save that subjects made membership judgements (as in Experiment 6) and four response options were introduced at stage 2. The task was amended as follows: as before, at stage 1, subjects had two response options. For example, "1. is a VEHICLE", "2. is not a VEHICLE". At stage 2 four response options for the conjunction were given. For example, "1. is a VEHICLE WHICH IS ALSO A MACHINE", "2. is just a VEHICLE", "3. is just a MACHINE" and "4. is neither a VEHICLE nor a MACHINE". Subjects indicated their responses by circling the appropriate number.

Subjects

10 subjects participated in the experiment. All subjects were students in full time education in Edinburgh, and participated in the experiment as part of their course work. Two subjects were excluded from the experiment as they were not native English speakers. The sexes were approximately equally represented in the group,

and subjects were naive to the purpose of the experiment.

RESULTS AND DISCUSSION

Giving the subjects more choice at stage 2 did not alter the number of overextensions (9.4%) and underextensions (3.4%) from Experiment 6 (12% and 3.1%, respectively). (Mann-Whitney U tests between Experiments 6 & 7 showed no significant differences for overextensions ($U = 19.5$, $df = 8$) or underextensions ($U = 30.5$, $df = 8$)). This result demonstrates that subjects' tendency to overextend cannot be reduced by providing more response options at stage 2.

Subjects' tendency to overextend appears to be robust under the manipulation of response option. This finding provides further evidence that overextensions are not caused by response biases. Further, Hampton (1988) reports that manipulation of the proportion of members and non-members, and therefore the proportion of positive and negative responses made by subjects, seems to have little influence on the tendency to overextend. These findings strongly suggest that overextension is a result of the processes of conceptual combination.

Summary

To summarize, Hampton's non-Boolean results were replicated in Experiment 4 (overextensions 15.8%, underextensions 2.9%). When subjects were asked to make typicality judgements for the same items on a 7 point scale, in Experiment 5, a hypothetical category boundary at the midpoint of the scale suggested that subjects "overextend" to a similar degree when making typicality judgements alone (overextensions 14.6%, underextensions 1.2%) The tendency to overextend was found to be robust when subjects were only asked to make membership judgements (overextensions 10.2%, underextensions 3.1%). These results were significantly different from those predicted by a Boolean model, as shown by comparison with the control condition of Experiment 6 (overextensions 5.3%, underextensions 4.2%). The residual difference of around 6% overextensions did not diminish when a wider range of response options were given in Experiment 7.

4.6. Intensional Combination and Theories

In Chapter 4 membership judgements in noun-noun combinations have been considered in some detail, and it has been clearly demonstrated that the Boolean model of conceptual combination is inadequate. The question that now has to be answered is whether this suggests that concepts are non-compositional. A number of authors are proposing that combination cannot be solved extensionally, which is why the Boolean model fails, but that concepts may still be intensionally compositional. I will consider below whether an intensional model of conceptual combination that can fare better than a Boolean account.

Perhaps the most explicit account of how intensional combination might work has been put forward in the "Inheritance of Attributes" model (Hampton, 1987). It is appropriate that this model be considered, and an assessment made of its appropriateness in explaining category combination, since the tendency for subjects to overextend was first documented by Hampton. In the inheritance of attributes model it is proposed that the intension of a conjunction is formed as the *union* of the intensions of the constituent concepts. As a first step, this predicts the same result as a Boolean model. That is, if the membership conditions for PET and the membership conditions for FISH are taken together, then any item fulfilling them would be a PET FISH. However, instead of assuming that intensions consist of defining features (as a Boolean model does) it is proposed that attributes (features) have a certain "importance" value for the category. It is predicted that the importance of an attribute for the conjunction will rise predictably as its importance in each of the constituent categories rises.

When an attribute is important for one constituent category, but not the other, it is predicted that the attribute will not be "inherited" by the conjunction. This is because the attribute has fallen below an average critical importance for the compound concept. Similarly, if the attributes of the constituent concepts conflict (e.g. PET might be said to have an attribute <animate>, LOG might be said to have the attribute <inanimate>, when the concepts combine to form PET LOG, there is an attribute conflict) only one attribute can be inherited, if the concept is to be coherent. (Here again coherence is invoked as a constraint in category formation - see Section 3.1). As yet, there are no proposals about how this conflict is resolved. Although it is not argued explicitly, presumably, it is the failure of attributes to be inherited that could be used to account for overextension effects. Thus, an item which was judged to be a non-member of a constituent category because it did not satisfy the requirements of a particular attribute, could become a legitimate member of the conjunct category if that attribute was not inherited. This would be classified as an overextension.

A qualification to the notion of inheritance failure is that subjects' intuitions about necessity and impossibility are always inherited. For example, subjects might think it necessary that PETS <have an owner>, while this is not an attribute of FISH. So, in the conjunction PET FISH the necessary attribute <has an owner> would be inherited. Similarly, subjects might think it impossible that FISH are <warm and cuddly>, although they might think that an important attribute for PETS, hence, there would be an inheritance failure of the impossible attribute in PET FISH.

Finally, it is proposed that complex concepts can acquire new attributes through a process of "extensional feedback", where experience in, and of, the world causes an alteration in a concept. This process can provide an explanation for the fact that in the Stroop Experiment, complex and complicated forms of sentence containing the words "scavenger" and "bird" had different meanings - see Appendix 2, Sectiion 2. This mechanism could also be used to account for the phenomenon of underextension. Thus, an item could be judged a member of both constituent categories but not a member of the conjunction, because the conjunction has an additional attribute, the requirements of which, the item does not fulfil. It seems to me, that the notion of extensional feedback could only work if the conjunctive concept were "lexicalized". In this case, a new attribute could be learned and incorporated into the concept. However, it is highly unlikely that many of the numerous possibilities of conceptual combinations are lexicalized. In these cases, it is unclear that there would be processing time available for extensional feedback to produce a new attribute in the intension of the compound, before the category decisions and inferences the compound was being used for, had already been resolved by some other method. Thus, I argue that extensional feedback is only a useful proposal for explaining underextensions in lexicalized compounds. Hampton acknowledges something of this argument when he puts forward the prediction that extensional feedback will make familiar compounds less compositional than unfamiliar ones.

Hampton conducted several experiments to test his model of intensional combination. He asked a number of subjects to generate attribute lists for 12 categories. Having obtained a reliable list of attributes for each concept, he asked a second group of subjects to assess the importance of the attributes for each constituent category and both forms of the conjunction (A PET which is also a FISH, and, A FISH which is also a PET). The inheritance of attributes model predicts that an item's importance for the constituent categories is a predictor of its importance for the conjunction. If a count is made from the average importance ratings of attributes (see Table 11) it is found that there are 47 cases where an attribute with a high average importance for both constituents also had a high average importance for the conjunction. Whereas there are only 4 cases where an attribute rated important for both constituents was not inherited by one form of the conjunct, and 1 case where it was not inherited by either form. This confirms the model's prediction that importance in the constituent is a good predictor of importance in the conjunct. The number of attributes thought important for one (but not the other) constituent which were thought important for both forms of the conjunction was 42, those attributes only inherited by one form of the conjunction numbered 10, while those not inherited by either form were 13. This lack of attribute inheritance could be used to account for overextensions.

Table 11

Number of attributes whose average importance (I) > 1.5, for constituent categories and conjunctions. Attributes with an importance rating equal to 1.5 were excluded from the count. (This data was compiled from the average importance ratings for attributes in the appendix of Hampton, 1987).

	I in Both Conjunctions	I in Only One Conjunction	I in Neither Conjunction
I in Both A and B	47	4	1
I in A or B	42	10	13
I in neither A nor B	4	6	27

There were 10 cases where the average importance for an item was low in both constituent categories, but high for at least one form of the conjunct. Thus, compound categories can inherit "unimportant" attributes from their constituents, or acquire new "important" attributes. This is consonant with the fact that subjects make underextensions. However, as no theoretical explanation is given of *why* unimportant constituent attributes become important for the conjunction (save for extensional feedback, but see above) this model does not provide an *explanation* of underextension.

Hampton found that two other factors were involved in attribute inheritance. Firstly, the head-qualifier effect was found to alter the relative contributions of constituents to the conjunction. Thus, it seems that the concept "A PET which is also a FISH" has a different intension from the concept "A FISH which is also a PET". Secondly, concept dominance was an important factor in determining attribute inheritance. That is, some concepts were found to be more important than others in forming the intension of the conjunct. There is some suggestion that concept dominance may result from having more important attributes than other concepts. The head-qualifier effect and concept dominance effects interact when concepts combine. Thus, the inheritance of attributes model provides an account of how concepts might combine interactively. It proposes a weighted average importance for each attribute, with unimportant attributes usually being excluded from the conjunction. A supplementary process of extensional feedback may provide a means for compound concepts to acquire new attributes under certain conditions.

The important thing to note about this account, is that the crucial determiners of how concepts combine are the weightings on particular attributes. Moreover, these weightings are not stable values, but change in different contexts e.g. with the linguistic context of the head-modifier relation. It is also likely, that which attribute is inherited when there is a conflict will be a result of the context. For example, Hampton suggests that TOOL may have <is used for construction> as an attribute, while WEAPON has <is used for destruction>. He argues that the destructive aspect is more central, since tools can be used constructively or destructively, and is hence more likely to be inherited in the compound A TOOL which is also a WEAPON (or A WEAPON which is also a TOOL). While this may well be true, which attribute is actually inherited in a given situation will be context dependent. Thus, in the context: "The soldier used his weapon to dig an entrance to the tunnel", the attribute <is used for construction> would be more likely to be inherited.

This leads me to the conclusion that the explanation for interactive conceptual combination is not provided in the intensions of concepts *per se*, but in the perceived relations between concepts in a particular context. In the inheritance of attributes model, the real explanation lies in the change of feature weightings. As yet, the reasons behind these changes have not been discovered. In Chapter 5 I will consider an alternative explanation of overextension, and whether it can yield any testable predictions about how concepts combine.

CHAPTER 5

WHY OVEREXTEND: COMPENSATION BETWEEN DIMENSIONS?

5.1. Introduction [2]

Chapter 4 showed that when subjects judge membership in conjunctive categories they are biased in favour of overextending their categorizations. This finding raises the question of why subjects overextend? In section 4.6 the inheritance of attributes model of overextension was considered in some detail. It was noted that an explanation of why subjects overextend, pitched at a higher level, was needed. As I have argued elsewhere (Lyon & Chater, in press) a possible explanation of overextension comes from an understanding of the way in which people judge membership of conjunctive categories in everyday life. It seems that subjects are more lenient in their membership judgements the more factors they have to take into account. For example, suppose that you have the task of casting the lead role in a local play. You are looking for a tall, handsome, blond male with a convincing Scottish accent, and a retentive memory. Given that such people are rather rare you will, of course, compromise on certain dimensions - you might settle for a sandy haired man of medium height. This amounts to overextending the category. Previously you would not have considered sandy hair an instance of the category blond, nor medium height as tall; yet now you include this man in the compound category "a tall, handsome, blond male with a convincing Scottish accent". You would be especially likely to do this if he fulfilled the rest of your criteria.

I will refer to this explanation of overextension as the Compensation Hypothesis. Such compensation effects can be seen in many uses of property conjunctions: for example, when recognizing a stranger from a description (tall, dark and handsome with a red carnation, under the clock in Paddington Station). Another example is deciding what property fits the category "A house I would buy". When buying a house a person may have a long list of criteria which must be simultaneously

---

[2]The experiments reported in this chapter were carried out in collaboration with Nicholas Chater, Centre for Cognitive Science, University of Edinburgh. Material contained in this chapter is also being published (Lyon & Chater, in press).

fulfilled. For instance: affordable price, two bedrooms, separate kitchen, central location, garage, garden. Since there are so few houses for sale which meet all of these criteria, most people have to compromise. It is interesting to note that there is a tendency not to admit that the initial criteria have not been met. Instead, people extend their categories to incorporate instances not previously included. For example, what would previously have been described as a high price is now construed as an affordable price (by borrowing up to the limit!). Similarly, a concrete patio would not initially be described as a garden. If the house fulfils the other criteria sufficiently well, a concrete patio may satisfy the demand for a garden, as the buyer's category judgement becomes more lenient. In fact, buyers are often adamant that they have fulfilled all their criteria. In a process of rationalization they may argue that with house prices going up they wanted to take on a large mortgage, and that a patio provides an excellent, easily maintained garden with the addition of a few potted plants.

The tendency for people to alter their views to eliminate inconsistency, has been widely recognized in social psychology (particularly, Festinger, 1957). When faced with an instance that does not fulfil all of the desired criteria people may do one of three things: 1) they can give up some of their criteria (they did not *really* want two bedrooms); 2) they can admit that the instance is not the perfect house, and does not meet all their criteria; 3) they can avoid either of these unpalatable possibilities by widening their categories, so that their chosen house fulfils them all. My suggestion is that this third option is the source of the tendency to overextend conjunctive categories. Thus there is a motive for overextension in many social categorization judgements. However, it is highly unlikely that such motives are involved when making decisions such as whether or not "chess" is a member of the conjunctive category "a sport which is also a game". What has been documented is that people are highly motivated to order instances whatever the circumstances:

> People have in fact a fateful predilection for linear orderings. They are prone to place elements in linear orderings to the exclusion

of other structures and they handle linear orderings more facilely than most other structures.

De Soto, London & Handel, 1965

Given such a prediliction, a possible motive for overextension in this case might be the desire to ascribe a substantial proportion of the list to be rated to the compound category. To the contrary this, and other, task features have been shown in previous experiments not to be the cause of subjects overextending (Hampton, 1988; Chapter 4). I propose the explanation for overextension is simply that a subject's classification strategy is learned through everyday experience of the world. As such experience includes making categorization judgements for cases which do motivate the subject to overextend (in order to elliminate uncomfortable inconsistencies) overextension becomes a generally effective strategy when judging membership in complex categories.

The Compensation Hypothesis asserts that the more criteria an item must simultaneously fulfil, the more leniently the subject will judge the item's membership of the category. This leads to the prediction: the more constituent categories form a compound the greater the subject's tendency to overextend.

5.2. Experiment 8

The Compensation Hypothesis

This experiment is designed to test the Compensation Hypothesis by examining the way subjects conjoin three categories. This experiment is an adaptation of Experiment 6 where the rationale for the control condition was as follows: at stage 1, all subjects gave categorization judgements for single categories; at stage 2, the experimental group gave judgements for the conjuctions, while the control group repeated stage 1. The control group's responses at stage 2 were then conjoined using a Boolean strategy. This provided an estimate of the overextensions and underextensions resulting from forgetting and changing one's mind. Hence providing a direct estimate of spurious non-Boolean effects. In a similar way, appropriate controls for the conjunction of three categories were included in Experiment 8. These controls enabled the number of overextensions for triple conjunctions to be compared with the number of overextensions for binary conjunctions.

According to the Compensation Hypothesis, the more constituent categories joined together, the more category judgements should be overextended. Thus, judging membership of the triple conjunctive category "A FOOD, A PLANT AND A FLAVOURING", should lead to more overextensions than judging membership of any two of its binary conjunctions (e.g. "A FOOD AND A FLAVOURING").

METHOD

Design

The design is a 2 X 3 mixed design. The within subjects factor is response type (underextensions and overextensions). The between subjects factor is experimental group (1, 2 or 3). Within a conjunct, the order of the constituents was counterbalanced across subjects, such that 2 subjects rated each of the 6 possible orders in groups 2 and 3. Otherwise the design was identical to that of Experiment 4.

Subjects

36 subjects took part in the experiment, 12 in each of 3 groups. Subjects were students resident in Edinburgh, and participated in the experiment voluntarily.

Materials

Stimuli were adapted from the items and categories used in the previous experiments. A third category was added to every pair of categories. Six category triples were used: Machine - Vehicle - Toy, Furniture - Household Appliance - Luxury, Pet - Bird - Predator, Building - Dwelling - Business, Food - Plant - Flavouring, Weapon - Tool - Farm Equipment. The parenthetical "or part of a plant" used in the previous experiments (following Hampton) was omitted as unwieldy. A slight difference in syntax was also employed. Extending the previous syntactic form (A WEAPON WHICH IS ALSO A TOOL) would have been cumbersome (i.e. A WEAPON WHICH IS ALSO A TOOL WHICH IS ALSO FARM EQUIPMENT). Hence, the simpler form A WEAPON AND A TOOL was adopted for the binary case, and A WEAPON, A TOOL AND FARM EQUIPMENT for triple conjunctions. A list of 16 items was adapted from the previous list such that approximately equal numbers of items likely to produce each kind of response were included (that is 2 items considered to be likely members of all categories (+++); 2 items considered to be likely members of the first two categories (++-) and so on, for all 8 possibilities). Thus, the materials were such that roughly half of the items were judged to be members of each single category (although, as noted above, Hampton (1988) found that this proportion does not seem to effect membership judgements). Items were typed in random order and headed by the appropriate category name, e.g. A Machine; A Vehicle; A Toy; A Machine and A Vehicle; A Vehicle and A Machine; A Machine and A Toy; A Toy and A Machine; A Vehicle and A Toy; A Toy and A

Vehicle; A Machine, A Vehicle and A Toy; A Machine, A Toy and A Vehicle; A Vehicle, A Toy and A Machine; A Vehicle, A Machine and A Toy; A Toy, A Machine and A Vehicle; A Toy, A Vehicle and A Machine. Lists of all possible permutations were prepared in order to counterbalance for order effects.

The lists and categories used in Experiment 8 overlap significantly with those in Experiments 4-7. However, since new categories and items have been introduced, and thus a different proportion of items will be judged to be members of each category, the numbers of overextensions and underextensions are not directly comparable with the experiments reported in Chapter 4.

Procedure

Each of the three groups made single category judgements at stage 1 (e.g. A FOOD; A FLAVOURING; A PLANT). The design of the experiment at stage 2 is summarized in Figure 2.

FIGURE 2

Task Summary for Experiment 8

	Stage 1	Stage 2
Group 1	A B C	A B C
Group 2	A B C	(A B) C
Group 3	A B C	(A B C)

Categories within brackets are conjoined by the subjects; all other conjunctions are Boolean and are performed by the experimenter. Note that, in Group 2, (A B) denotes any possible pair.

Group 1

Group 1 served as a control group, and thus at stage 2 received single category judgements as they had at stage 1. The stage 2 judgements for each category triple were Booleanly conjoined, as in Experiment 6. For example, at stage 1, a subject might judge "tomato" to be a FOOD, a PLANT but not a FLAVOURING (++-). At the second stage they might judge "tomato" to be a member of all three categories (+++). The Boolean conjunction of the responses at stage 2 is "+", whereas the Boolean conjuction of the responses at stage 1 is "-". This is an overextension. This control provides an estimate of the number of overextensions and underextensions expected, even if the subject always makes Boolean judgements. Not all changes between responses at stage 1 and stage 2 will result in an overextension or an underextension. Suppose that a subject responds (+-+) at stage 1 and (++-) at stage 2. The Boolean conjunction of the responses at stage 2 is "-". This is consistent with the Boolean conjunction of the stage 1 judgement.

Group 2

At stage 2, Group 2 combined two categories of the triple, e.g. A FOOD AND A FLAVOURING, and judged the remaining category, A PLANT, separately. Equal numbers of lists were headed by each possible pair of categories (in this case, Food-Plant, Food-Flavouring, Plant-Flavouring). Some subjects judged the conjunct A FOOD AND A PLANT and the single category A FLAVOURING, some judged the conjunct A FLAVOURING AND A PLANT and the single category FOOD, and so on. Thus the results are not biased by choosing one particular pair of categories.

The judgement for the conjunct (e.g. +) and the judgement for the single category (e.g. -) were then Booleanly combined (-) to provide an estimate of the number of overextensions and underextensions resulting from the combination of two categories by the subjects. It is necessary to have subjects judge the single category so that the results of all three groups are directly comparable. This is because some of the overextensions in the binary cases would not be picked up in the triple conjunction.

Group 3

At stage 2, Group 3 judged the membership of the triple conjunction (e.g. A FOOD, A PLANT AND A FLAVOURING). In all other ways the procedure was the same as that of Experiment 6.

Rationale for Controls

If a comparison is to be made between groups to assess non-Boolean effects, it is crucial that the task is such that a strictly Boolean model of conjunction predicts the same number of overextensions and underextensions in each group. Hence, any difference in the number of overextensions or underextensions between groups is evidence for non-Boolean combination. This consideration led to the inclusion of single category judgements (for the remaining category of the triple) in addition to binary judgements, at stage 2. Thus, subjects judged, for example, the binary category A TOOL AND A WEAPON and also the single category FARM EQUIPMENT. It may seem more natural to have subjects judge the binary conjunction alone and compare the number of overextensions and underextensions directly with those judging triple conjunctions in Group 3. However, if this strategy were adopted a purely Boolean model would predict a greater number of overextensions and underextensions in Group 2 than in Group 3. This is because some overextensions and underextensions will be counted in Group 2, but hidden in Group 3. Suppose that a subject uses a completely Boolean strategy. In this case all overextensions and underextensions are due to forgetting, changing one's mind and so on. At stage 1, this subject might judge a "hammer" to be A TOOL, but not A WEAPON or FARM EQUIPMENT (+--). At stage 2, the subject might judge a "hammer" A TOOL and A WEAPON but not FARM EQUIPMENT. We now show, that using the attractive method of counting overextensions and underextensions described above, leads to this perfectly Boolean subject being classified as an overextender if in Group 2, but not in Group 3. If the subject is in Group 2, "hammer" will be judged to be A TOOL AND A WEAPON at stage 2, though it was not judged to be a WEAPON at stage 1. Hence, the subject would have overextended. On the other hand, the same change of mind does not lead to an overextension in Group 3. "Hammer" is rated as both A TOOL and A WEAPON at stage 2, but since it is not FARM EQUIPMENT, "hammer" is judged not to be A

TOOL, A WEAPON AND FARM EQUIPMENT. This is illustrated in Figure 2.

In order that overextensions in Groups 1,2 & 3 be comparable, it is necessary to incorporate a judgement about the third category (FARM EQUIPMENT) in all groups. Hence, for Group 2, it is necessary to compute the Boolean conjunction of judgements for the binary category (A WEAPON AND A TOOL) and the single category (FARM EQUIPMENT). The outcome of this procedure (in this case "–") is then compared to the subject's stage 1 judgements to assess the number of overextensions and underextensions appropriately. On a Boolean model of conjunction, the number of overextensions and underextensions should be the same across groups – see Figure 3.

The Compensation Hypothesis predicts that the more categories a subject conjoins the greater the number of overextensions will be made. Thus, there will be more overextensions in Group 3 than in Group 2, and more in Group 2 than in Group 1. In contrast, as I have noted, the Boolean hypothesis predicts that there should be no difference in the number of overextensions or underextensions in the three groups. Furthermore, considerations analagous to those outlined above show that, the number of overextensions should equal the number of underextensions in each group.

FIGURE 3

		A WEAPON	A TOOL	FARM EQUIPMENT
Group 2				
	Stage 1	+	-	-
	Stage 2	(+	+)	-

Stage 2 (++) is Booleanly conjoined to give (+) for A WEAPON AND A TOOL. This is an overextension.

		A WEAPON	A TOOL	FARM EQUIPMENT
Group 3				
	Stage 1	+	-	-
	Stage 2	(+	+	-)

Stage 2 (++-) is Booleanly conjoined to give (-) for A WEAPON A TOOL AND FARM EQUIPMENT. This is not an overextension.

A 2 X 3 mixed ANOVA was performed. A main effect of response type was found ($F(1, 33) = 16.7$, $p < 0.05$). Inspection of the means showed that this was due to subjects' tendency to overextend rather than underextend their categories (see Table 12). The main effect of group was also significant ($F(2,33) = 9.72$, $p < 0.05$). There is a significant interaction between response type (overextensions vs underextensions) and group (single judgements, binary judgements, triple judgements) ($F(2,33) = 6.62$, $p < 0.05$). Post hoc t-tests showed that, as expected, there was no difference between the number of overextensions and underextensions in Group 1 ($t = 0.07$ (11), $p > 0.05$) (See Table 12). In Groups 2 and 3, however, there were significantly more overextensions than underextensions ($t = 3.33$ (11), $p < 0.01$ in Group 2; $t = 3.56$ (11), $p < 0.01$ in Group 3). As argued above, if subjects used a Boolean combination strategy, the number of overextensions and underextensions should be the same for all groups. A post-hoc Joncqheere Trend test showed that there was a significant increase in overextensions from Group 1 through to Group 3 (S=160 (12), $p < 0.05$). These results show that the more constituents the subjects have to conjoin, the greater their propensity to overextend. This finding is exactly what the Compensation Hypothesis predicts.

Table 12

Overextensions and underextensions made in Experiment 8, as a function of the number of categories subjects conjoined.

	Overextensions	Underextensions
Group 1 (single judgements)	34	33
Group 2 (binary conjunction)	53	19
Group 3 (triple conjunction)	117	27

5.3. Implications of a Compensation Model

I have proposed that conjunctive categorization judgements are a result of a best fit strategy. Thus, when subjects have a number of criteria to meet simultaneously, they tend to interpret these criteria more leniently than when they are judged independently. So when a simple category is in the context of being a constituent of a complex category, it is more likely to be overextended by the subject. Concept combination can be construed as a particular sort of context effect where each constituent of the compound acts as a context for the other. For example, in understanding the complex concept PET FISH, the word "pet" influences the kind of fish we expect, and the word "fish" influences the kind of pet we expect, because of what we know about fish and pets. In this light the phenomenon of overextension can be construed as another example of the context dependency of category judgements.

Contextual influence in concept combination has been explained as a matter of making best fit judgements to a number of criteria. This view may be extended to the general problem of contextual influence on categorization. Categorization judgements may be seen as "best fit" matches to the constraints context has imposed. In such a view, just as the word "pet" provides a context which alters what is judged to be a fish, the real life context of a pet shop alters what is judged to be a fish.

If concept combination is a product of best fit matching, it will not be possible to understand when and why particular concepts may be overextended without adverting to world knowledge. Overextension, and conceptual flexiblity, is constrained by general knowledge about the situation. Thinking back to the example of choosing someone for a theatrical role, it may be allowable to construe 5'8" as tall, and an ordinary looking man as handsome. On the other hand, in the context of a play about Edinburgh it may not be possible to overextend the criterion that the actor has a Scottish accent. This is because a Scottish audience would find such a compromise unacceptable. However, this requirement too is subject to context. If the production is touring Japan then an Irish accent might do just as well. The compromises made in best fit matching appear to be a product of understanding the immediate situation. Thus, in order to have a full account of conceptual combination, it may be necessary to take account of subjects' understanding of the world. A compensation account of concept combination is thus consonant with the view that concepts must be understood in the context of theories of world (e.g. Murphy & Medin, 1985).

Murphy & Medin (1985) and Lakoff (1986) have proposed an alternative to the compositional account put forward in the inheritance of attributes model. Instead they argue that concepts are not compositional in any simple sense. These authors point out that conceptual features are not independent, but interrelated. Hence, we know that fish are cold and slimy *because* we know they live in water, and so on. Medin and others have proposed that conceptual combination is a process of discovering, or inventing, a relation between concepts. He emphasizes the importance of seeing concepts as embedded in theories. A theory based account of conceptual combination can also explain the data accounted for in the inheritance of attributes model. For instance, if people use relational theories to organize meaning, then it might be expected that an attribute listing experiment would not reflect the intensions of concepts, but the implications of the particular theory the person had in mind. Thus, it is not intensions which combine when words join, but different theories which are invoked, and which concentrate on different relations.

Another cause of overextensions may be what Tversky & Kahneman (1983) have called "the conjunction fallacy", when the generation of a theory links (perhaps fallaciously) two previously unrelated concepts. Tversky & Kahneman have demonstrated that people rate two unlikely events which are related together in a theory as more likely to occur than either event independently. For example, people thought it unlikely that "A massive flood somewhere in North America in

1983, in which more than 1,000 people would drown", unlikely to occur. The same people thought it more likely, however, that "An earthquake in California sometime in 1983" would cause "a flood in which more than 1,000 people drown". This could be described as an overextension: flood (–), earthquake (+), flood and earthquake (+).

The problem with a theory approach to concepts is that the approach has not been clearly articulated. It is a theoretical notion which seems to be saying that in order to understand any word we need to employ all our world knowledge. Moreover, rather than that world knowledge being organized in concepts, as previously thought, it seems that it is unstructured and we must access all or nothing! In Chapter 6, I will argue that knowledge can be represented in a way consonant with a theory approach. The idea that concepts are the result of theories is, I believe, a major step forward in trying to address the real questions of why concepts combine in the way they do.

A possible objection to using the facts of overextension as support for the appropriateness of a theory based approach, might be that Hampton's Inheritance of Attributes model (see section 4.6) is also capable of explaining the increase in overextension with triple conjunctions. Such an explanation would be that as more categories are introduced, more attributes conflict. When these conflicts are resolved in the conjunction, items previously excluded from the constituent category are admitted to membership in the conjunction. What is not clear is what effect increasing the number of categories to be conjoined would have on other response options. For example, it might be that a new category introduces new attributes which were not required for any of the other constituents, this may lead to an increase in the number of consistent responses, by excluding previous overextensions (see Figure 3). Since this is possibly the case, it is unclear that an Inheritance of Attributes model *predicts* a substantial increase in overextensions in Experiment 8. This is because in a study of 16 items the introduction of a third category is just as likely to increase the number of additional new attributes (leading to more consistent responses, and thus excluding possibilities to overextend) as additional attribute conflicts (leading to overextensions). Thus, although the Inheritance of Attributes model can provide an *ad hoc* explanation of the increase in overextensions with triple conjunctions, it does not clearly *predict* such a change.

Moreover, the Compensation Hypothesis has several theoretical advantages over an Inheritance of Attributes model. Firstly, it does not require the decomposition of each category into attributes which are subsequently combined according to a set of combination rules. As has been noted elsewhere, such a proposition has serious drawbacks (see Sections 4.6 and 6.2). Secondly, the Compensation Hypothesis has the advantage of not having to construe overextensions as by-products of processing, but has a theoretical framework in which an investigation of their strategic value is appropriate. As I have argued previously it may be more productive to view context effects as "normal" categorization rather than as complicated exceptions. Similarly, overextension may not be most profitably investigated as an exception to "normal" logical combination. Both phenomena are central to our categorization processes, and may well be highly motivated. The Compensation Hypothesis provides a framework in which it is appropriate to investigate what may be an important classification strategy.

CHAPTER 6

CATEGORIES AND CONCEPTS: A FUTURE PERSPECTIVE

6.1. Conspectus

The aim of this thesis has been to investigate the structure of concepts, the effect of contextual change on that structure and, in particular, the process of conceptual combination.

In Chapter 1, four functions that concepts are used to fulfil are described. Firstly, concepts are used to explain stability in response to particular stimuli, and thus to enable effective communication. Secondly, concepts are thought to represent ontological and linguistic information which enables simple inference. Thirdly, concepts are thought to contain the information people use to make simple categorization decisions, to be an intrinsic part of how those decisions are made. Fourthly, concepts are similarly used to explain complex categorizations.

Three major theories of concepts were assessed with respect to these four conceptual functions. It was noted that the Classical View could not provide an adequate explanation of categorization phenomena, because of its reliance on defining features. Thus, the need for the inclusion of characteristic features in the conceptual representation was highlighted. This amended position is close to the Probabilistic Views which provide a more satisfactory account of simple categorization. However, the Probabilistic Views find it difficult to explain complex categorization, particularly since Fuzzy Set Theory has been heavily criticized. The third theory of concepts considered was the Exemplar View. This theory differs from the others, by rejecting the notion of a summary category concept, and insisting on a holistic representation. The Exemplar View provides a good account of simple categorization phenomena, but seems to imply that concepts are completely non-compositional. It was noted that the adoption of a holistic representation excludes the concept from fulfilling any linguistic/ontological role. The first chapter concluded with a discussion of the possible implications of categorization on conceptual theories, in light of the debate on epistemology versus metaphysics.

In Experiment 1, the question of conceptual stability was addressed. An assessment was made of the changes in GOE as a result of context. Changes in context were found to completely restructure the GOE distribution. Differences in sentence structure were found to have little effect on meaning, supporting the notion that complex concepts are routinely formed in contexts where the qualifier does not occur immediately prior to the noun. Chapter 2 concludes with an assessment of the usefulness of the notion of similarity in explaining conceptual structure. The close relationship between similarity and conceptual coherence is noted, although it is argued that similarity alone is not sufficient to explain coherence. The question was raised of how context relevant attributes, over which similarity can be computed, are selected.

The issue of conceptual coherence was explored further in Chapter 3. It was concluded that coherence may be one measure people use to constrain meaning. The question of whether concepts can provide conceptual coherence was assessed, by examining the range of context effects on an on-line reading task. Experiment 2, demonstrated that context is immediately taken into account when making categorization decisions while reading. It was found that in the context of a target sentence (e.g "You found it!" said Andy, noticing the broken tool in the workshop) modified picture probes (broken hammer) are identified as members of the category (tool) more quickly than more typical, unmodified pictures (hammer). The possibility that this was the result of subjects using an overall matching strategy, in which the adjective and the noun were matched to the picture independently, was assessed in Experiment 3. In this case, the adjective was not represented in the picture independently from the noun (e.g. fairground fruit). The conclusion from Experiment 3, was that the effect of context could not be explained simply in terms of an overall matching strategy, but that context influenced the *retrieval* of category exemplars. The fact that context effects cannot be ascribed to a post-access period, make it unlikely that coherence is a result of conceptual representation alone. Medin's suggestion that coherence results from theories is considered in this light. The idea, that the process by which relevant attributes are selected in different contexts results from the formation of different complex concepts, was proposed. An understanding of the process of concept combination was seen as vital for the development of conceptual theory.

In Chapter 4, the process of conceptual combination was studied. Although it is the prevailing view of conceptual combination that concepts combine compositionally, it was demonstrated that subjects do not form conjunctions according to a Boolean model. Rather, it was found that subjects overextend their categories; that is, they include items in conjuncts they have previously excluded from constituent categories. This tendency was found to be robust under several manipulations. Since the Boolean account of conceptual combination seemed inadequate, an alternative account, proposing intensional compositionality, was examined. Although an intensional account seemed to hold more promise in describing conceptual combination, and explaining overextension, it was noted that most of the explanatory work was being done by changes in feature weightings, that were peripheral to the theory. It was concluded that the explanation for the interactive combination of concepts, could not be provided in the intensions of concepts alone, but rather, must include the relationships between concepts in a particular context.

In Chapter 5, an alternative account of overextension is considered. It is suggested that people have a motive to overextend in many everyday categorization decisions, and that this tendency is carried over to decisions where no such motive is obvious. Use of such a best fit strategy, predicts that the larger the categories conjoined, the greater the tendency to overextend. In Experiment 8, subjects judging membership of triple conjuncts overextended considerably more than those judging membership of binary conjuncts. In considering the implications of this effect, the Inheritance of Attributes model was reconsidered, but the importance of adopting a theory based approach was reiterated. It was noted that a possible problem with such an approach, is the conception of how theories might be represented.

In this chapter I will conclude that most conceptual theories have a very limited applicability. I will argue that this is so, because of the assumption that concepts have localized representations. An argument for a globalist view of representation is presented, and it is demonstrated that it is consonant with Medin's theory based approach. I conclude by outlining how localist and globalist approaches might be distinguished experimentally.

6.2. Localism and Globalism [3]

With the exception of a theory based approach to meaning, current theories of concepts all hold the implicit theoretical assumption that concepts are psychologically localised. That is, a concept corresponds to a unit which *contains* the information necessary to understand that concept (Collins & Quillian, 1969; Glass & Holyoak, 1975; Katz & Postal, 1964; Smith, Shoben & Rips, 1974; Miller & Johnson-Laird, 1976; McCloskey & Glucksberg, 1979; Fodor, 1981; Barsalou, 1982). For instance, many theorists ascribe to each lexical item of natural language an entry in a mental lexicon, which must be accessed to understand the word. Thus, to have a concept is to have the appropriate 'packet' of information. Intuitively it seems that the meaning of a phrase (e.g. "black cat") is typically derived from the meaning of its parts (the words "black" and "cat"). Correspondingly, complex concepts (e.g. BLACK CAT) are derived from their constituent concepts (BLACK and CAT). According to the localist view, concept combination is a matter of putting simple packets together. I argued in Section 4:1 that many theorists hold such a compositional view of meaning, although there appears to be no satisfactory account of how constituent prototypes can be combined compositionally to produce complex prototypes (although see section 4.6 for a contender). In order to develop a theory which can adequately account for the flexibility and context sensitivity of prototypes, and prototype combination, the implicit localism of current views may have to be rejected.

<u>Why is a New Account Needed?</u>

In chapters 2 and 3 it was demonstrated that prototype views fail to provide an inkling of how concepts may be combined. Hence, it seems unacceptable to identify concepts with prototype representations. To capture compositionality it appears necessary to invoke the Definitional view of the structure of concepts, and

---

[3] Material contained in this section is currently being published (Lyon & Chater, in press)

to explain prototype effects as side effects of processing rather than as functions of the representation *per se*. It was demonstrated in Chapter 4 that a Definitional account, which seems able to capture the apparent compositionality of phrases such "black cat", cannot provide an adequate explanation of membership decisions for simple conjunctions. Thus, it was concluded that a Definitional account alone cannot provide a complete theory of concepts.

One reason why it may have been so hard to develop a theory of concepts capable of handling concept combination is the implicit localism of many current theories. That is, the problem of concept combination is thought to be that of providing a function which somehow composes the packets of information corresponding to the constituents into a packet of information corresponding to the complex concept. However, as I have claimed, the problem of concept combination may in fact be a special case of the more general phenomenon of contextual influence on concepts. If this is the case then just as context effects are generally agreed to depend upon world knowledge, world knowledge may be crucially implicated in concept combination. I shall consider below evidence against the localist notion of concepts. I will argue that no localist account of concepts can be maintained, and propose that a globalist alternative, in which prototype effects are knowledge driven expectations, must be developed. According to a globalist view, concepts are a function of high level knowledge of the world, computed on the fly.

<u>Problems with Localism</u>

A localist view is intuitively appealing - when we say that a child has the concept DOG it sounds as if that child possesses some fixed mental structure (the concept DOG), which the child accesses when dogs are recognized. Once the structure is in place, it can be used to construct memories about dogs, generate and understand sentences involving the word "dog" and so on. Thus the localised concept provides a building block for the mental operations of recognition, memory, and language. However, the localist view has difficulty in accounting for a wide range of well-known data. In particular it is hard to explain the facts that (i) concepts are highly flexible and context sensitive; and that (ii) as has just been shown, concepts do not appear to combine compositionally. I propose that it is the localised notion of concepts which makes these ubiquitous phenomena so difficult to explain.

In order to explain prototype effects some notion of typicality, or similarity to a prototype is used. This is problematic for the localist because the prototype and the notion of similarity used appear to be a function of the context. Thus, at the fishmonger my prototypical fish might be close to a cod; and my similarity judgements might be based on flavour, appearance and price. On the other hand, at the pet shop my prototypical fish might be near a goldfish; and my similarity metric might be based on size, shape and living conditions. There only seem to be two explanations for this phenomenon which appear to be available to the localist. The first is that these uses of the term 'fish' are simply polysemous: that is that these uses correspond to distinct lexical entries in the same way as (river) bank and (high street) bank. That lexical items have a finite set of distinct senses in the mental lexicon is the "sense selection assumption" (Clark, 1983). However, polysemy is an inadequate explanation of context sensitivity since there appear to be as many goodness of fit distributions as there are contexts (Roth & Shoben, 1983). Further it appears that prototypicality effects can be obtained for completely novel concepts like people from London with green hair (Barsalou, 1983). It seems most plausible that rather than retrieving stored goodness of fit distributions from memory we create them on line. Indeed, the goodness of example distribution seems to be a reflection of what the subjects expect an instance to be like, given their knowledge of the world. Hence prototypes are a function of the whole cognitive process, rather than being building blocks out of which such processes can be constructed. It is simply the recognition of these facts that distinguishes the globalist view.

The other explanation of the flexibility and context dependence of prototypicality, open to the localist, is that changes in the goodness of example distribution are a result of post-access inferences. That is, the same conceptual representation is retrieved on all occasions of use, but subsequent inferences then determine the contextual plausibility of an instance (Gumenik, 1979; Whitney & Kellas, 1984). As reported in Chapter 2, Whitney & Kellas (1984) claimed to provide empirical support for this point of view, by using a Stroop paradigm. They found, for example, that "bird" will prime the word "robin" more than the word "chicken", even in a context such as "the guest saw the bird that roasted on the grill", which strongly suggests the less typical bird. However, I did not replicate these results in a similar experiment reported in Chapter 2. Subsequent experiments reported in Chapter 3 (after Potter & Faulconer, 1979) appear to show immediate contextual modification in a picture recognition task, given a sentential context. This implies that contextual modification is immediate and not reliant on post access reasoning.

It is a central task for a theory of concepts to explain how it is that inference and world knowledge can influence the goodness of example distribution in response to context. So, even if there is an underlying localist mechanism of combination which is not sensitive to context (which is an unresolved question), a complete explanation of the data will require a detailed consideration of world knowledge rich effects. Until there is firm evidence to the contrary, parsimony surely dictates that a 'two tier' explanation of prototype phenomena is not postulated.

A Globalist Alternative

Murphy & Medin (1985) have pointed out the role of naive theories of the world in conceptual coherence. Even in the early prototype views (Rosch, 1973) it was postulated that the prototype is a function of the way in which the structure of the world is described within a particular culture. This shared cultural knowledge was held to determine the prototype that an individual must have in order to possess a concept. Thus the theories by which the world is understood are crucially implicated in determining the structure of prototypes. According to this picture, world knowledge is distilled into a localisable packet, the fixed prototype. This approach must be extended, however, to account for subjects' rapid construction of contextually appropriate goodness of example distributions. For example, if the words "musical instrument" are heard at a Classical concert hall my prototype might be close to violin; if I am going to a rock concert my prototype might be close to electric guitar.

The primary goal of a theory of concepts must be to explain the everyday use of words in real contexts. Even in cases where context is underspecified (e.g. hearing the words "musical instrument" as the radio is switched on) people attempt to provide their own contexts using their common sense theories about what makes most sense in the circumstances (for example, what radio station it is). Thus, context effects are the rule rather than the exception. Context effects cannot be viewed as limited deformations of some fixed, stored, conceptual representation which applies in the alleged 'null' context.

In this thesis, I have demonstrated that neither Prototype nor Definitional views appear to provide an adequate explanation of context effects. Rather than localising the representation of the meaning of a word in some packet of information, the process of understanding seems more likely to be a matter of the generation of expectations on the basis of our world knowledge, rather than assessing similarity to a prototype. Consider the analogy of a landmine. States of the world can be 'categorized' as: those in which the mine detonates, and those in which it does not. Thus, the mine can be considered as a categorizor - dividing the world into two. The states of the world in which the mine detonates (distant earthquakes, nearby tanks, floods causing water to leak in, electric storms, and so on) need have nothing in common other than that they detonate the mine. Further, which states detonate the mine is a function of the structure of the whole mine (how shock resistant, waterproof, and resistant to electrical activity it is). The 'categorization' of the world into detonating and non-detonating circumstances cannot be explained by saying that the mine possesses some localisable structure which corresponds to the 'concept' DETONATE (Note 9). Analogously, human categorization judgements cannot be understood without considering the cognitive system at large. Thus, I propose that concepts cannot be identified as psychologically localised structures but are globally determined by the cognitive system.

The problem of deploying our world knowledge appropriately has been found to be crucially implicated in a wide variety of cognitive domains (Minsky, 1975). Some workers in Artificial Intelligence have attempted to understand the way in which our theories of the world are represented using fixed primitive concepts as building blocks (Schank & Abelson, 1977; Hayes, 1978). Perhaps one reason why such approaches have been so unsuccessful (Fox & McDermott, 1986) is that they treat concepts as rigid blocks, where in fact they are as mysterious, unstable, and flexible as the processes they are used to model. Current successes of PDP models may be due in part to their utilization of the principle of distributed representation.

A globalist theory of concepts may seem alarming since it seems that to understand concepts our common sense theories of the world must also be understood. However, difficult problems cannot be solved by providing simple solutions to easy problems. The theory approach to concepts as proposed by Medin and others, seems compatible with a globalist approach to concepts. I believe it is only with this type of theory that real progress can be made in understanding conceptual representation.

6.3. A Developmental Perspective on Localism and Globalism

In the previous section I proposed a globalist hypothesis in which concepts are characterized as being generated on line as a function of the whole cognitive system (see also Lyon & Chater, in press). Globalism emphasizes the constructive nature of concepts, and the central role of world knowledge in such constructions. This view is consonant with a change of emphasis currently advocated by some researchers (e.g. Douglas Medin, Ulric Neisser, George Lakoff, Benny Shanon). This position is radically different from traditional "localist" theories of representation which rely on the retrieval of "stored packets of meaning". Context effects and conceptual combination have proved stumbling blocks for all localist theories, this may indicate that a conceptual jump, such as adopting a globalist view, is needed to provide the necessary apparatus to tackle these phenomena. Although the dispute between localist and globalist theories of representation is a high level one, the approaches do predict differences with regard to concept acquisition.

According to the globalist view, the possession of a concept is a direct function of the ability to apply world knowledge. In this way, categorization is directly tied to the theory of the domain in which the categorization is being made. Thus, globalism predicts that the concept a child uses in a particular situation will depend on the theory the child has of that situation. This yields an important prediction for newly acquired concepts. This prediction can be illustrated with the example of a child learning the concept FISH. In the fish shop, the child's concept is based on the kind of fish usually found there: cod fillets, whole trout, smoked kippers, shellfish. So, in this context, the child learns to categorize fish using criteria appropriate to the domain (such as fillets vs whole fish, smell, colour of flesh, taste etc). For the globalist, even at the earliest stages of development, concepts will be context sensitive, as the concepts used to make a categorization are *created* relative to the domain.

The globalist predicts that to be able to extract uniformity across contexts is more difficult than to understand a concept in a particular context. Thus, the abstract theories which draw out the similarities between words in different contexts should be a late cognitive development. The realization that fish in the fish shop share many properties with fish in the pet shop, for example, should only arise when the child's theory is sufficiently well developed. For the child to use some of the same criteria to judge fish in the fish shop and fish in the pet shop, demands that the child's theory must be broad enough to relate together pet fish and fish for tea. Thus, the globalist predicts that to be able to extract uniformity across contexts is more difficult than to understand a concept in a given context. Since context sensitivity is the norm, and stability across contexts is a late development, the globalist prediction is that children's categorization judgements will be *more* context dependent than those of adults.

The globalist prediction is in direct contradiction to more traditional localist accounts; which assume that to possess a concept is to possess a stored packet of meaning. The localist perceives concepts as simple building blocks, and context effects as producing complicated modifications of these basic units. In the same way, complex concepts are characterized as the result of a process which puts simple packets together. The localist explains the context sensitivity and flexibility of concepts by claiming that the concept is first accessed, then modified in a contextually appropriate way. Thus, the application of world knowledge to produce appropriate contextual changes is seen as a high level addition to the basic categorization process. Hence, the ability to manipulate these stored concepts, producing conceptual flexibility, should gradually evolve during development. Although it should be expected that children's concepts are extremely variable over time, and between individuals, the localist view suggests that children's concepts will not be as flexible in response to context as those of adults. Thus, the localist prediction is that children's categorization judgements will be *less* context sensitive than those of adults.

The differing predictions of localist and globalist theories with respect to concept acquisition, presents a welcome possibility of testing the appropriateness of a globalist view experimentally.

CONCLUSION

The issues of conceptual representation cannot hope to be solved in a thesis, and many aspects of the whole notion of representation are unsatisfactory. There are still no clear answers to how world knowledge is stored and utilized in thought and language. Investigating the notion that world knowledge is stored in concepts has led me to the conclusion that large amounts of world knowledge must be fundamentally involved in the interpretation of every "concept". The fact that theory approaches concentrate on the relational aspects of meaning, might make them better suited for modelling the highly complicated way meaning is ascribed to the world.

Notes

1. CAPITAL LETTERS are used to denote concepts. "Inverted commas" are used to denote words or phrases. Normal type is used to denote referents in the world.

2. Rey (1983) draws an important distinction between metaphysical and epistemological function which will be discussed later in the chapter (see section 1:5).

3. It may be noted that this approach makes no attempt to explain what it is that makes some instances more similar to the prototype than others. This is a just criticism but it is important to bear in mind that the ad hoc postulation of underlying features is not an explanation either, but the displacement of the problem to a lower level.

4. This point may be criticised by pointing out that concepts like APPLE and ORANGE are themselves abstracted from the total set of apples and oranges. A logical extension of the Exemplar View would simply state that the concept APPLE has a holistic prototypical representation consisting of a small group of best instances.

5. Weak versions of the Exemplar View simply allow for the representation of some abstract information in concepts. Theories of this form are subject to similar criticisms as the Probabilistic approaches.

6. There were two possible exceptions to this in the case of "pale vegetable" and "colourful bird". In these examples the "paleness" and "colourfulness" of the unmodified and modified pictures was matched. Hence any relative difference in response to modified probes could not be a result of how well the probe matched the adjective.

7. Potter & Faulconer's finding that modified probes did not have shorter response latencies than unmodified probes when the context sentence contained a modifying adjective separated from the noun, does not necessarily contradict Barsalou's notion. This is because Potter & Faulconer's task specifically encouraged subjects to ignore the sentence, while in real-life sentence processing, context is a central component.

8. Proof that the probability of overextensions and underextensions is equal. In Hampton's original formulae, there are separate parameters (s_1 and s_2) which denoted whether the category was rated first or second. In these experiments order is counterbalanced and hence no seperate parameters are necessary.

Underextensions:

$p(++-)$

$= s.g.[\ u^2.(0) + u.(1 - u).(1 - g) + (1 - u).u.(1 - s) + (1 - u)^2.(1 - s.g)\ ]$

$= s.g.[\ (u - u^2).(2 - s - g) + (1 - 2.u + u^2).(1 - s.g)\ ]$

$= s.g.[\ (u^2.(s + g - 1 - s.g)$

$\quad + u.(2.s.g - s - g)$

$\quad + \ (1 - s.g) \qquad]$

Overextensions:

$p(+-+) + p(-++) + p(--+)$

$= s.(1 - g).[u^2.(0) + u.(1 - u).g + (1 - u).u.(0) + (1 - u)^2.s.g]$

$+ (1 - s).g.[u^2.(0) + u.(1 - u).(0) + (1 - u).u.s + (1 - u)^2.s.g]$

$+ (1 - s).(1 - g).[(1 - u)^2.s.g]$

$= \qquad (u - u^2).[s.g.(1 - g) + s.g.(1 - s)]$

$+ (1 - 2.u + u^2).[s^2.g.(1 - g) + s.g^2.(1 - s) + s.g.(1 - s).(1 - g)]$

$= s.g.[\ (u^2.(s + g - 1 - s.g)$

$\quad + u.(2.s.g - s - g)$

$\quad + \ (1 - s.g) \qquad]$

9. Thanks to Nick Chater who provided this example.

References

Armstrong, S. L., Gleitman, L. R. and Gleitman, H. (1983) What Some Concepts Might Not Be. *Cognition*, **13**, 263-308.

Arnold, J. B. (1971) A Multidimensional Scaling Study of Semantic Distance. *Journal of Experimental Psychology (Monograph)*, **90**, 349-372.

Barsalou, L. W. (1982) Context-Independent and Context-Dependent Information in Concepts. *Memory and Cognition*, **10**, 82-93.

Barsalou, L. W. (1983) Ad Hoc Categories. *Memory and Cognition*, **11**, 211-227.

Barsalou, L. W. (1987) The Instability of Graded Structure: Implications for the Nature of Concepts. Chapter 5 in Neisser, U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization.* Cambridge: Cambridge University Press.

Beals, R., Krantz, D. H. and Tversky, A. (1968) Foundations of Multidimensional Scaling. *Psychological Review*, **75**, 127-142.

Beller, H. K. (1971) Priming: Effects of Advance Information on Matching. *Journal of Experimental Psychology*, **87**, 176-182.

Boyd, J. P. (1972) Information Distance for Discrete Structures. In *Multidimensional Scaling*, Volume 1: Theory: *Theory and Applications in the Behavioural Sciences*. New York: Seminar Press.

Brooks, L. R. (1987) Decentralized Control of Categorization: The Role of Prior Processing Episodes. Chapter 6 in Neisser, U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization.* Cambridge: Cambridge University Press.

Bruner, J. S., Goodnow, J. J. and Austin, G. A. (1956) *A Study of Thinking.* New York : Wiley.

Carey, S. (1982) Semantic Development: The State of the Art. Chapter 12 in Wanner, E. and Gleitman, H. (eds.) *Language Acquisition: The State of the Art.* Cambridge: Cambridge University Press.

Carroll, J. D. and Wish, M. (1974) Multidimensional Perceptual Models and Measurement Methods. In *Handbook of Perception.* New York: Academic Press.

Clark, E. V. (1973a) What's in a Word? On the Child's Acquisition of Semantics in his First Language. In *Cognitive Development and the Acquisition of Language.* New York: Academic Press.

Clark, E. V. (1973b) Non-linguistic Strategies and the Acquisition of Word Meanings. *Cognition*, **2**, 161-182.

Clark, H. H. (1983) Making Sense of Nonce Sense. Chapter 9 in d'Arcais, G. B. F. and Jarvella, R. J. (eds.) *The Process of Language Understanding*. Chichester: John Wiley and Sons.

Collins, A. M. and Quillian, M. R. (1969) Retrieval Time From Semantic Memory. *Journal of Verbal Learning and Verbal Behaviour*, **8**, 240-247.

Collins, A. M. and Loftus, E. F. (1975) A Spreading Activation Theory of Semantic Processing. *Psychological Review*, **82**, 407-428.

Conrad, C. (1974) Context Effects in Sentence Comprehension: A Study of the Subjective Lexicon. *Memory and Cognition*, **2**, 130-138.

Corbett, A. T. and Dosher, B. A. (1978) Instrumental Inferences in Sentence Encoding. *Journal of Verbal Learning and Verbal Behaviour*, **17**, 479-491.

De Soto, C. B. and London, M. and Handel, S. (1965) Social Reasoning and Spatial Paralogic. *Journal of Personal Social Psychology*, **2**, 513-521.

Dosher, B. A. and Corbett, A. T. (1982) Instrumental Inferences and Verb Schemata. *Memory and Cognition*, **7**, 240-244.

Dowty, D. R., Wall, R. E. and Peters, S. (1981) *Introduction to Montague Semantics*. Dordrecht: D. Reidel.

Egan, J. P. and Clarke, F. R. (1964) Source and Receiver in the Use of a Criterion. In Swets, J. A. (ed.) *Signal Detection and Recognition by Human Observers*. USA: Wiley.

Festinger, L. (1957) *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

Fodor, J. A. (1981) The Current Status of the Innateness Controversy. In *Representations*. Cambridge, Mass.: MIT Press.

Fox, M. S. and McDermott, J. (1986) The Role Of Databases in Knowledge-Based Systems. Technical Report No. CMU-RI-TR-86-3, The Robotics Institute, Pittsburgh, Pa., February, 1986.

Garner, W. R. (1978) Aspects of a Stimulus: Features, Dimensions and Configurations. In *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.

Garrod, S. and Sanford, A. J. (1977) Interpreting Anaphoric Relations: The Integration of Semantic Information While Reading. *Journal of Verbal Learning and Verbal Behaviour*, **16**, 77-90.

Glass, A. L. and Holyoak, K. J. (1975) Alternative Conceptions of Semantic Memory. *Cognition*, **3**, 313-339.

Goldman, D. and Homa, D. (1977) Integrative and Metric Properties of Abstracted Information as a Function of Category Discriminability, Instance Variability and Experience. *Journal of Experimental Psychology: Human Learning and Memory*, **3**, 375-385.

Gumenik, W. E. (1979) The Advantage of Specific Terms Over General Terms as Cues for Sentence Recall: Instantiation or Retrieval? *Memory and Cognition*, **7**, 240-244.

Hamilton, D. L. and Gifford, R. K. (1976) Illusory Correlation in Interpersonal Perception. A Cognitive Basis of Stereotypic Judgements. *Journal of Experimental Social Psychology*, **12**, 392-407.

Hampton, J. A. (1979) Polymorphous Concepts in Semantic Memory. *Journal of Verbal Learning and Verbal Behaviour*, **18**, 441-461.

Hampton, J. A. (1987) Inheritance of Attributes in Natural Concept Conjunctions. *Memory and Cognition*, **15**, 55-71.

Hampton, J. A. (1988) Overextension of Conjunctive Concepts: Evidence for a Unitary Model of Concept Typicality and Class Inclusion. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **14**, 12-32.

Hayes, P. (1978) The Naive Physics Manifesto. Working Paper No. 34, ISSCO Institut Dalle Molle, Geneva, Switzerland, 1978.

Hayes-Roth, B. and Hayes-Roth, F. (1977) Concept Learning and the Recognition and Classification of Exemplars. *Journal of Verbal Learning and Verbal Behaviour*, **16**, 119-136.

Henley, N. M. (1969) A Psychological Study of the Semantics of Animal Terms. *Journal of Verbal Learning and Verbal Behaviour*, **8**, 176-184.

Homa, D. and Silver, R. (1976) Triadic Decision Making in Lexical Memory. *Memory and Cognition*, **4**, 532-540.

Johnson-Laird, P. N. (1983) *Mental Models*. Cambridge: Cambridge University Press.

Kahneman, D. and Tversky, A. (1973) On the Psychology of Prediction. *Psychological Review*, **80**, 237-251.

Katz, J. J. and Postal, P. (1964) *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass: MIT Press.

Kendler, H. H. (1964) The Concept of a Concept. In Melton, A. W. (ed.) *Categories of Human Learning*. London: Academic Press.

Krantz, D. L. and Tversky, A. (1975) Similarity of Rectangles: An Analysis of Subjective Dimensions. *Journal of Mathematical Psychology*, **12**, 4-34.

Lakoff, G. P. (1973) Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. In Corum, C., Smith-Stark, T. C. and Weiser, A. (eds.) *Proceedings of the ninth annual meeting of the Chicago Linguistic Society*, Chicago, 1973.

Lakoff, G. P. (1986) *Women, Fire, and Dangerous Things: What Categories Tell us About the Nature of Thought*. Chicago : University of Chicago Press.

Lakoff, G. P. (1987) Cognitive Models and Prototype Theory. Chapter 4 in Neisser, U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge: Cambridge University Press.

Lyon, K. and Chater, N. (in press) Localist and Globalist Theories of Concepts. In Gilhooly, K. J., Keane, M., Logie, R. and Erdos, G. (eds.) *Lines of Thinking: Volume 1*. Chichester: J Wiley.

McCauley, R. N. (1987) The Role of Theories in a Theory of Concepts. Chapter 11 in Neisser, U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge: Cambridge University Press.

McCloskey, M. and Glucksberg, S. (1978) Natural Categories: Well defined or Fuzzy Sets?. *Memory and Cognition*, **6**, 462-472.

McCloskey, M. and Glucksberg, S. (1979) Decision Processes in Verifying Category Membership Statements. *Cognitive Psychology*, **11**, 1-37.

Medin, D. L. and Schaffer, M. M. (1978) Context Theory of Classification Learning. *Psychological Review*, **85**, 207-238.

Medin, D. L. and Smith, E. E. (1984) Concepts and Concept Formation. *Annual Review Psychology*, **35**, 113-138.

Medin, D. L. and Wattenmaker, W. D. (1987) Category Cohesiveness, Theories, and Cognitive Archeology. Chapter 3 in Neisser, U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorisation*. Cambridge: Cambridge University Press.

Miller, G. A. and Johnson-Laird, P. N. (1976) *Language and Perception*. Cambridge: Cambridge University Press.

Miller, G. A. (1978) Semantic Relations Among Words. In Halle, M., Bresnan, J. and Miller, G. A. (eds.) *Linguistic Theory and Psychological Reality*. MIT Press.

Minsky, M. (1975) A Framework for Representing Knowledge. Chapter 6 in Winston, P. (ed.) *The Psychology of Computer Vision*. New York: McGraw Hill.

Murphy, G. L. and Medin, D. L. (1985) The Role of Theories in Conceptual Coherence. *Psychological Review*, **92**, 289-316.

Neisser, U. (1987) Introduction: The Ecological and Intellectual Bases of Categorization. Chapter 1 in Neisser, U. (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge: Cambridge University Press.

Oden, G. C. (1977) Integration of Fuzzy Logical Information. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 565-575.

Osherson, D. N. and Smith, E. E. (1981) On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition*, **9**, 35-58.

Osherson, D. N. and Smith, E. E. (1982) Gradedness and Conceptual Combination. *Cognition*, **12**, 299-318.

Paivio, A. (1971) *Imagery and Verbal Processes.* New York: Holt, Rinehart and Winston.

Paris, S. and Lindauer, B. (1976) The Role of Inference in Children's Comprehension and Meaning for Sentences. *Cognitive Psychology*, **8**, 217-227.

Potter, M. C. and Faulconer, B. A. (1979) Understanding Noun Phrases. *Journal of Verbal Learning and Verbal Behaviour*, **18**, 509-521.

Quine, W. V. O. (1960) *Word and Object.* Cambridge, Mass: MIT Press.

Rey, G. (1983) Concepts and Stereotypes. *Cognition*, **15**, 237-262.

Richards, J., Platt, J. and Weber, H. (eds.) (1985) *Longman Dictionary of Applied Linguistics*. England: Longman, Chaucer Press.

Rips, L. J., Shoben, E. J. and Smith, E. E. (1973) Semantic Distance and the Verification of Semantic Relations. *Journal of Verbal Learning and Verbal Behaviour*, **12**, 1 - 20.

Rosch, E. H. (1973) On the Internal Structure of Perceptual and Semantic Categories. In *Cognitive Development and the Acquisition of Language*. New York: Academic Press.

Rosch (1975a) Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, **104**, 192-233.

Rosch, E. (1975b) Cognitive Reference Points. *Cognitive Psychology*, **7**, 532-547.

Rosch, E. (1978) *Principles of Categorization.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Rosch, E. and Mervis, C. B. (1975) Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, **7**, 573-605.

Rosch, E., Mervis, C., Gray, W., Johnson, D. and Boyes-Braem, P. (1976) Basic Objects in Natural Categories. *Cognitive Psychology*, **8**, 382-439.

Roth, E. M. and Shoben, E. J. (1983) The Effect of Context on the Structure of Categories. *Cognitive Psychology*, **15**, 346-378.

Roth, E. M. and Mervis, C. B. (1983) Fuzzy Set Theory and Class Inclusion Relations in Semantic Categories. *Journal of Verbal Learning and Verbal Behaviour*, **22**, 509-525.

Schank, R. C. and Abelson, R. (1977) *Scripts, Plans, Goals and Understanding*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Shepard, R. N. (1962) Scaling with an Unknown Distance Function: I. *Psychometrika*, **27**, 125-140.

Shepard, R. N. (1974) Representation of Structure in Similarity Data: Problems and Prospects. *Psychometrika*, **39**, 373-421.

Smith, E. E., Shoben, E. J. and Rips, L. J. (1974) Structure and Process in Semantic Memory: A Featural Model for Semantic Decisions. *Psychological Review*, **81**, 214-241.

Smith, E. E. and Medin, D. L. (1981) *Categories and Concepts*. Cambridge Mass: Harvard University Press.

Snodgras, J. G. and McClure, P. (1975) Storage and Retrieval Properties of Dual Codes for Pictures and Words in Recognition Memory. *Journal of Experimental Psychology: Human Learning and Memory*, **1**, 521-529.

Stroop, J. R. (1938) Factors Affecting Speed in Serial Verbal Reactions. *Psychological Monographs*, **50**, 38-48.

Tajfel, H. and Wilkes, A. L. (1963) Classification and Quantitative Judgement. *British Journal of Psychology*, **54**, 101-114.

Tversky, A. and Kahneman, D. (1983) Probability, Representativeness, and the Conjunction Fallacy. *Psychological Review*, **90**, 293-315.

Tversky, A. and Krantz, D. L. (1970) The Dimensional Representation and the Metric Structure of Similarity Data. *Journal of Mathematical Psychology*, **7**, 572-597.

Tversky, A. and Gati, I. (1978) Studies of Similarity. In Rosch, E. and Lloyd, B. (eds.) *Cognition and Categorization*. Hillsdale, New Jersey: Erlbaum.

Uyeda, K. M. and Mandler, G. (1980) Prototypicality Norms for 28 Semantic Categories. *Behavior Research Methods and Instrumentation*, **12**, 587-589.

Warren, R. E. (1972) Stimulus Encoding and Memory. *Journal of Experimental Psychology*, **94**, 90-100.

Whitney, P. and Kellas, G. (1984) Processing Category Terms in Context: Instantiation and the Structure of Semantic Categories. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **10**, 95-103.

Wittgenstein, L. (1953) *Philosophical Investigations*. Oxford: Basil Blackwell. Translated by G E M Anscombe.

Zadeh, L. A. (1965) Fuzzy sets. *Information and Control*, **8**, 338-353.

Zadeh, L. A. (1982) A Note on Prototype Theory and Fuzzy Sets. *Cognition*, **12**, 291-297.

APPENDIX 1

TABLE A1: FREQUENCY LISTS OF EXAMPLES FOR THREE FORMS OF TEN TYPICAL SENTENCES.

	EXEMPLAR	FREQ	%	EXEMPLAR	FREQ	%

The woman screamed "Help" as she stared at the *weapon* in his hand.

	EXEMPLAR	FREQ	%	EXEMPLAR	FREQ	%
FIRST	Gun	13	56.5	Sword	1	4.3
	Knife	6	26.0	Blade	1	4.3
	Revolver	1	4.3	Sawn-Off Shotgun	1	4.3
SECOND	Knife	9	39.1	Dagger	2	8.7
	Gun	7	30.4	Pistol	1	4.3
	Axe	2	8.7	Long knife	1	4.3

The woman screamed "It's *bloody*" as she stared at the *weapon* in his hand.

	EXEMPLAR	FREQ	%	EXEMPLAR	FREQ	%
FIRST	Knife	12	54.5	Gun	1	4.5
	Dagger	3	13.6	Machete	1	4.5
	Sword	3	13.6	Bread Knife	1	4.5
SECOND	Knife	4	18.2	Pitchfork	1	4.5
	Axe	4	18.2	Gun	1	4.5
	Hammer	3	13.6	Crowbar	1	4.5
	Club	3	13.6	Razor	1	4.5
	Sword	2	9.1	Hatchet	1	4.5
	Chainsaw	1	4.5			

The woman screamed "Help" as she stared at the *bloody weapon* in his hand.

	EXEMPLAR	FREQ	%	EXEMPLAR	FREQ	%
FIRST	Knife	11	50	Axe	3	13.6
	Dagger	7	31.8	Scalpel	1	4.5
SECOND	Axe	5	22.7	Truncheon	1	4.5
	Sword	3	13.6	Gun	1	4.5
	Knife	2	9.1	Poker	1	4.5
	Blade	2	9.1	Club	1	4.5
	Dagger	2	9.1	Broken Bottle	1	4.5
	Chainsaw	1	4.5	Spear	1	4.5
	Cudgel	1	4.5			

Though Sam had it with him he hoped that the *weapon* wouldn't be necessary.

FIRST	Gun	15	71.4	Switchblade	1	4.8
	Knife	4	19	Penknife	1	4.8
SECOND	Knife	7	33	Rifle	1	4.8
	Dagger	2	9.5	Revolver	1	4.8
	Gun	2	9.5	Pipe	1	4.8
	Pistol	2	9.5	Grenade	1	4.8
	Cudgel	1	4.8	Baton	1	4.8
	Flick knife	1	4.8	Truncheon	1	4.8

Though Sam had it with him in a *harness* he hoped that the *weapon* wouldn't be necessary.

FIRST	Gun	9	42.9	Bow	1	4.8
	Machine gun	3	14.3	Cosh	1	4.8
	Crossbow	3	14.3	Sawn-off shotgun	1	4.8
	Pistol	1	4.8	Tank	1	4.8
	Lunger	1	4.8			
SECOND	Sword	5	26.3	Revolver	1	4.8
	Knife	4	19.0	Machete	1	4.8
	Rifle	2	9.5	Catapult	1	4.8
	Harpoon Gun	1	4.8	Anti-aircraft gun	1	4.8
	Flame thrower	1	4.8	Dagger	1	4.8
	Bazooka	1	4.8			

Though Sam had it with him he hoped that the *harnessed weapon* wouldn't be necessary.

FIRST	Gun	8	36.4	Bow & Arrow	1	4.5
	Crossbow	3	13.6	Dagger	1	4.5
	Rifle	2	9.1	Pistol	1	4.5
	Missile	2	9.1	Sword	1	4.5
	Machine gun	2	9.1	Hatchet	1	4.5
SECOND	Knife	6	31.6	Truncheon	1	5
	Gun	3	15	Rifle	1	5
	Bomb	2	10	Bow & Arrow	1	5
	Pistol	2	10	Sword	1	5
	Crossbow	1	5	Icepick	1	5
	Cannon	1	5			

Attempting to complete the job, Dick moved the *vehicle* across the site.

FIRST	Lorry / Truck	8	36.4	Car	1	4.5
	Bulldozer	6	27.3	Earthmover	1	4.5
	J.C.B.	2	9.1	Tractor	1	4.5
	Excavator	1	4.5	Digger	1	4.5
	Dumper Truck	1	4.5			
SECOND	Lorry / Truck	5	22.7	Combine Harvester	1	4.5
	Tractor	4	18.2	Wheelbarrow	1	4.5
	Dumper	3	13.6	Fork-Lift Truck	1	4.5
	Van	2	9.1	Car	1	4.5
	Crane	2	9.1	Caterpillar Truck	1	4.5
	Bulldozer	1	4.5			

Attempting to complete the job, Dick *loaded* and moved the *vehicle* across the site.

FIRST	Truck / Lorry	11	52.4	Dump Truck	1	4.8
	Fork-Lift Truck	3	14.3	J.C.B.	1	4.8
	Van	2	9.5	Digger	1	4.8
	Car	2	9.5			
SECOND	Van	6	28.6	Tractor	2	9.5
	Lorry / Truck	6	28.6	Fork-Lift Truck	2	9.5
	Car	2	9.5	Bike	1	4.8
	Crane	2	9.5			

Attempting to complete the job, Dick moved the *loaded vehicle* across the site.

FIRST	Lorry / Truck	8	38.1	Trailer	1	4.8
	Van	4	19.0	Pick-Up Truck	1	4.8
	Dump Truck	3	14.3	Car	1	4.8
	Wheelbarrow	2	9.5	Bulldozer	1	4.8
SECOND	Van	5	23.8	Cart	1	4.8
	Truck / Lorry	4	19.0	Bus	1	4.8
	Wheelbarrow	3	14.3	Articulated Lorry	1	4.8
	Tractor	3	14.3	J.C.B.	1	4.8
	Digger	2	9.5			

"It's over there" directed Mark, hoping Harry would find the *tool* where he left it.

FIRST	Hammer	12	52.2	Chisel	1	4.3
	Screwdriver	3	13.0	Drill	1	4.3

	Spade	3	13.0	Saw	1	4.3
	Spanner	2	8.7			
SECOND	Hammer	8	34.8	Saw	2	8.7
	Screwdriver	6	26.1	Pliers	1	4.3
	Chisel	3	13.0	Spanner	1	4.3
	Drill	2	8.7			

"It's *hanging* over there" directed Mark, hoping Harry would find the *tool* where he left it.

FIRST	Hammer	10	45.5	Screwdriver	2	9.1
	Saw	6	27.3	Hoe	1	4.5
	Spanner	2	9.1	Spade	1	4.5
SECOND	Hammer	9	40.9	Hedge Clippers	1	4.5
	Screwdriver	3	13.6	Chainsaw	1	4.5
	Saw	4	18.2	Garden Fork	1	4.5
	Spanner	2	9.1	Hatchet	1	4.5

"It's over there" directed Mark, hoping Harry would find the *hanging tool* where he left it.

FIRST	Hammer	9	40.9	Rake	1	4.5
	Saw	7	31.8	Spanner	1	4.5
	Spade	2	9.1	Grass Cutters	1	4.5
	Hoe	1	4.5			
SECOND	Saw	8	36.4	Screwdriver	1	4.5
	Hammer	2	9.1	Fork	1	4.5
	Spade	2	9.1	Wooden Spoon	1	4.5
	Axe	2	9.1	Spanner	1	4.5
	Scythe	1	4.5	Drill	1	4.5
	Shovel	1	4.5	Shears	1	4.5

"You found it!" said Andy, noticing the *tool* in the workshop.

FIRST	Hammer	10	43.5	Saw	1	4.3
	Spanner	3	13.0	Screwdriver	1	4.3
	Lathe	2	8.7	Spade	1	4.3
	Chisel	2	8.7	Plane	1	4.3
	Drill	2	8.7			
SECOND	Saw	4	18.2	Spanner	1	4.5
	Screwdriver	4	18.2	Car Jack	1	4.5
	Chisel	3	13.6	Axe	1	4.5

	Fork	2	9.1	Chainsaw	1	4.5
	Hammer	2	9.1	Stanley Knife	1	4.5
	Plane	2	9.1			

"You *broke* it!" said Andy, noticing the *tool* in the workshop.

FIRST	Screwdriver	4	18.2	Drill	1	4.5
	Hammer	3	13.6	Plane	1	4.5
	Spanner	3	13.6	Spirit-Level	1	4.5
	Saw	2	9.1	Vice	1	4.5
	Chisel	2	9.1	Axe	1	4.5
	Hacksaw	2	9.1	Hatchet	1	4.5
SECOND	Hammer	3	14.3	Plane	1	4.8
	Hacksaw	2	9.5	Drill	1	4.8
	Screwdriver	2	9.5	Clamp	1	4.8
	Saw	2	9.5	Workmate	1	4.8
	Chisel	2	9.5	Spanner	1	4.8
	Chainsaw	1	4.8	Rivet Gun	1	4.8
	Set Square	1	4.8	Axe	1	4.8
	Power drill	1	4.8			

"You found it!" said Andy, noticing the *broken tool* in the workshop.

FIRST	Hammer	6	27.3	Saw	1	4.5
	Screwdriver	5	22.7	Screw	1	4.5
	Spanner	2	9.1	Spade	1	4.5
	Chisel	2	9.1	Lathe	1	4.5
	Scythe	1	4.5	Corkscrew	1	4.5
	Shovel	1	4.5			
SECOND	Saw	5	22.7	Nail	1	4.5
	Hammer	4	18.2	Fork	1	4.5
	Chisel	4	18.2	Spanner	1	4.5
	Screwdriver	3	13.6	Spade	1	4.5
	Mallet	1	4.5			

Helen put it aside because she was in a hurry, even though the *fruit* appealed to her.

FIRST	Banana	7	31.8	Pineapple	1	4.5
	Peach	6	27.3	Grapefruit	1	4.5
	Apple	4	18.2	Passion Fruit	1	4.5
	Orange	2	9.1			
SECOND	Apple	8	38.1	Nectarine	1	4.8
	Banana	5	23.8	Orange	1	4.8

	Grapes	2	9.5	Kiwi Fruit	1	4.8
	Pear	1	4.8	Lychee	1	4.8
	Pineapple	1	4.8			

Helen put it aside *half-eaten* even though the *fruit* appealed to her.

FIRST	Apple	15	68.2	Orange	1	4.5
	Pear	3	13.6	Melon	1	4.5
	Peach	1	4.5	Banana	1	4.5
SECOND	Banana	5	22.7	Peach	2	9.1
	Apple	5	22.7	Strawberry	1	4.5
	Orange	5	22.7	Grapefruit	1	4.5
	Pear	3	13.6			

Helen put it aside because she was in a hurry, even though the *half-eaten fruit* appealed to her.

FIRST	Apple	16	72.7	Banana	1	4.5
	Orange	2	9.1	Pineapple	1	4.5
	Pear	2	9.1			
SECOND	Orange	6	27.3	Grapefruit	1	4.5
	Banana	5	22.7	Pear	1	4.5
	Peach	3	13.6	Strawberry	1	4.5
	Apple	3	13.6	Plum	1	4.5
	Apricot	1	4.5			

"They are all there" said Linda, as her mother viewed the *fruit* in the hall.

FIRST	Banana	6	26.1	Mixed Fruit	1	4.3
	Orange	5	21.7	Grapes	1	4.3
	Apple	5	21.7	Grapefruit	1	4.3
	Pear	3	13.0	Pineapple	1	4.3
SECOND	Banana	5	22.7	Pineapple	2	9.1
	Grapes	2	9.1	Coconut	1	4.5
	Apple	2	9.1	Pawpaw	1	4.5
	Orange	2	9.1	Melon	1	4.5
	Grapefruit	2	9.1	Lemon	1	4.5
	Pear	2	9.1	Peach	1	4.5

"They are all *piled*" said Linda, as her mother viewed the *fruit* in the hall.

FIRST	Apple	9	40.9	Melon	2	9.1
	Orange	6	27.3	Pear	1	4.5
	Banana	3	13.6	Mango	1	4.5
SECOND	Orange	6	28.6	Lemon	1	4.8
	Apple	4	19.0	Pear	1	4.8
	Pineapple	2	9.5	Cherries	1	4.8
	Grapes	2	9.5	Guava	1	4.8
	Peach	1	4.8	Grapefruit	1	4.8
	Banana	1	4.8			

"They are all there" said Linda, as her mother viewed the *piled fruit* in the hall.

FIRST	Orange	6	27.3	Coconut	2	9.1
	Apple	5	22.7	Melon	1	4.5
	Banana	4	18.2	Grapefruit	1	4.5
	Grapes	2	9.1	Sharon Fruit	1	4.5
SECOND	Apple	8	36.4	Watermelon	1	4.5
	Orange	3	13.6	Grapefruit	1	4.5
	Pineapple	2	9.1	Tomato	1	4.5
	Coconut	2	9.1	Grapes	1	4.5
	Banana	2	9.1	Cherries	1	4.5

Although it was on loan, Kathy used the piece of *furniture* all the time.

FIRST	Table	6	28.6	Wheelchair	1	4.8
	Chair	4	19.0	Hairbrush	1	4.8
	Desk	2	9.5	Hi-fi	1	4.8
	Sofa	2	9.5	T.V.	1	4.8
	Armchair	2	9.5	Bed	1	4.8
SECOND	Chair	6	28.6	Desk	1	4.8
	Table	5	23.8	Mirror	1	4.8
	T.V.	2	9.5	Arm Chair	1	4.8
	Piano	2	9.5	Sofa	1	4.8
	Bed	1	4.8	Cooker	1	4.8

Although it was an *antique*, Kathy used the piece of *furniture* all the time.

FIRST	Chair	8	34.8	Chest of Drawers	1	4.3
	Table	3	13.0	Writing Desk	1	4.3
	Piano	2	8.7	Armchair	1	4.3
	Desk	2	8.7	Crockery	1	4.3
	Rocking Chair	1	4.3	Chamber Pot	1	4.3
	Wardrobe	1	4.3	Porcelain Jug	1	4.3

SECOND	Table	7	30.4	Upright Chair	1	4.3
	Desk	3	13.0	Bed	1	4.3
	Chair	3	13.0	Sideboard	1	4.3
	Bureau	2	8.7	Silverware	1	4.3
	Cupboard	2	8.7	Chaise Longue	1	4.3
	Sofa	1	4.3			

Although it was on loan, Kathy used the piece of *antique furniture* all the time.

FIRST	Chair	5	25	Bookcase	1	5
	Chest of Drawers	3	15	Dressing Table	1	5
	Table	3	15	Tea Chest	1	5
	Desk	2	10	Chest	1	5
	Dresser	1	5	Mirror	1	5
	Couch	1	5			
SECOND	Table	6	30	Bed	1	5
	Chest of Drawers	3	15	Lamp	1	5
	Chair	3	15	Sideboard	1	5
	Spinning Wheel	1	5	Coffee Table	1	5
	Oil Lamp	1	5	Wardrobe	1	5
	Piano	1	5			

The teenager had already acquired some small articles of *furniture* of her own.

FIRST	Table	4	33.3	Welsh Dresser	1	8.3
	Chair	1	8.3	Lamp	1	8.3
	Painting	1	8.3	Bedside Table	1	8.3
	Dressing Table	1	8.3	Bean Bag	1	8.3
	Desk	1	8.3			
SECOND	Chair	3	25	Stereo Cabinet	1	8.3
	Table	2	16.7	Stool	1	8.3
	Bean Bag	2	16.7	Water Bed	1	8.3
	Cutlery	1	8.3	Bookshelf	1	8.3

The teenager had already acquired some small articles of *furniture* of *modern* design.

FIRST	Chair	3	37.5	Table	1	12.5
	Lamp	2	25	Coffee Table	1	12.5
	Desk	1	12.5			
SECOND	Chair	2	25	Table	1	12.5
	Bed	2	25	Lamp	1	12.5
	Shelves	2	25			

The teenager had already acquired some small articles of *modern furniture* of her own.

FIRST	Chair	3	23.1	Couch	1	7.7
	Lamp	3	23.1	Wardrobe	1	7.7
	Coffee Table	1	7.7	Bean Bag	1	7.7
	Table	1	7.7	Hi-fi	1	7.7
	Stool	1	7.7			
SECOND	Lamp	3	23.1	Table	1	7.7
	Desk	2	15.4	Chair	1	7.7
	Sofa Bed	1	7.7	Bookcase	1	7.7
	Bean Bag	1	7.7	Stool	1	7.7
	Bedside Table	1	7.7	Video	1	7.7

Although it was borrowed, Jill thought that the *clothing* would be adequate.

FIRST	Coat	6	27.3	Jumper	1	4.5
	Dress	4	18.2	Ball Gown	1	4.5
	Waterproofs	3	13.6	Skirt	1	4.5
	Trousers	2	9.1	Tracksuit	1	4.5
	Jacket	1	4.5	Wedding Dress	1	4.5
	Dungarees	1	4.5			
SECOND	Ski Suit	2	9.1	Habit	1	4.8
	Trousers	2	9.1	Jacket	1	4.8
	Hat	2	9.1	Overall	1	4.8
	Jumper	2	9.1	Skirt	1	4.8
	Dress	1	4.8	Coat	1	4.8
	Nightdress	1	4.8	Gown	1	4.8
	Blouse	1	4.8	Scarf	1	4.8
	Shoes	1	4.8	Suit	1	4.8
	Boots	1	4.8	Jacket	1	4.8

Although it was old, Jill thought that the *clothing* would be adequate.

FIRST	Dress	4	18.2	Thermal Vest	1	4.5
	Coat	3	13.6	Anorak	1	4.5
	Jumper	3	13.6	Cloak	1	4.5
	Ball Gown	2	9.1	Fur Coat	1	4.5
	Jacket	2	9.1	Shawl	1	4.5
	Bikini	1	4.5	Suit	1	4.5
	Jeans	1	4.5			
SECOND	Coat	7	31.8	Hat	2	9.5
	Jumper	4	18.2	G-string	1	4.5
	Dress	2	9.5	Fur Coat	1	4.5
	Skirt	2	9.5	Shirt	1	4.5
	Boots	2	9.5			

Although it was borrowed, Jill thought that the *old clothing* would be adequate.

FIRST	Dress	6	26.1	Dungarees	1	4.3
	Jumper	4	17.4	Trousers	1	4.3
	Shawl	2	8.7	Jacket	1	4.3
	Cardigan	2	8.7	Skirt	1	4.3
	Coat	2	8.7	T-shirt	1	4.3
	Raincoat	1	4.3	Hat	1	4.3
SECOND	Skirt	4	17.4	Boned Blouse	1	4.3
	Trousers	3	13.0	Raincoat	1	4.3
	Shirt	3	13.0	Dress	1	4.3
	Coat	2	8.7	Hat	1	4.3
	Cardigan	2	8.7	Jacket	1	4.3
	Jumper	1	4.3	Climbing Boots	1	4.3
	Jeans	1	4.3	Cloak	1	4.3

APPENDIX 1 Continued

TABLE A2: FREQUENCY LISTS OF EXAMPLES FOR THREE FORMS OF TEN ATYPICAL SENTENCES.

	EXEMPLAR	FREQ	%	EXEMPLAR	FREQ	%

The state authorities used their chosen *weapon* to execute criminals.

FIRST	Electric Chair	5	25	Political Pressure	1	5
	Hanging	5	25	Hobnail Boots	1	5
	Rifle	2	10	Machine Gun	1	5
	Gun	2	10	Noose	1	5
	Guillotine	2	10			
SECOND	Electric Chair	5	25	Thumbscrew	1	5
	Guillotine	2	10	Gas Chamber	1	5
	Noose	2	10	Firing Squad	1	5
	Rifle	2	10	Shooting	1	5
	Hanging	1	5	Axe	1	5
	Gun	1	5	Drug Injection	1	5
	Pill	1	5			

The state authorities used their chosen *weapon* to execute criminals *humanely*.

FIRST	Electric Chair	7	41.2	Gun	1	5.9
	Rifle	3	17.6	Guillotine	1	5.9
	Hanging	2	11.8	Injection	1	5.9
	Revolver	1	5.9	Gas	1	5.9
SECOND	Gun	7	41.2	Guillotine	1	5.9
	Electric Chair	3	17.6	Gas	1	5.9
	Needle	2	11.8	Poison	1	5.9
	Hanging	2	11.8			

The state authorities used their *humane weapon* to execute criminals.

FIRST	Electric Chair	5	33.3	Chainsaw	1	6.7
	Guillotine	3	20	Axe	1	6.7
	Injection	2	13.3	Execution Method	1	6.7
	Gas	1	6.7	Noose	1	6.7

SECOND	Injection	2	15.4	Gas Chamber	1	7.7
	Drugs	2	15.4	Gun	1	7.7
	Anaesthetic	2	15.4	Rope	1	7.7
	Guillotine	2	15.4	Electric Chair	1	7.7

The young rioter spotted a *weapon* in the street and hurled it at the policeman.

FIRST	Brick	5	29.4	Bomb	1	5.9
	Stone	5	29.4	Hand Grenade	1	5.9
	Petrol Bomb	4	23.5	Bottle	1	5.9
SECOND	Bottle	6	33.3	Tomato	1	5.6
	Brick	3	16.7	Petrol Bomb	1	5.6
	Grenade	2	11.1	Tyre	1	5.6
	Flick Knife	1	5.6			

Conveniently, the young rioter spotted a *weapon* in the street and hurled it at the policeman.

FIRST	Brick	11	64.7	Stone	1	5.9
	Bottle	2	11.8	Projectile	1	5.9
	Dustbin Lid	1	5.9	Scaffolding Clip	1	5.9
SECOND	Bottle	5	31.2	Plank	1	6.2
	Brick	3	18.8	Fence Post	1	6.2
	Stick	2	12.5	Pick	1	6.2
	Stone	1	6.2	Gun	1	6.2

The young rioter spotted a *convenient weapon* in the street and hurled it at the policeman.

FIRST	Brick	8	40	Rock	1	5
	Bottle	7	35	Metal Bar	1	5
	Stone	2	10	Pipe	1	5
SECOND	Brick	7	35	Stick	1	5
	Bottle	3	15	Iron Bar	1	5
	Milk Bottle	2	10	Can	1	5
	Stone	1	5	Piece of Wood	1	5
	Boulder	1	5	Chain	1	5
	Dustbin Lid	1	5			

Nick thought that the *bird* he saw at the zoo was very tame.

FIRST	Penguin	3	12.5	Macaw	1	4.2
	Eagle	3	12.5	Canary	1	4.2
	Ostrich	3	12.5	Cockatoo	1	4.2
	Robin	2	8.3	Buzzard	1	4.2
	Emu	2	8.3	Budgie	1	4.2
	Owl	2	8.3	Minah Bird	1	4.2
	Parakeet	1	4.2	Parrot	1	4.2
	Flamingo	1	4.2			
SECOND	Parrot	5	22.7	Robin	1	4.5
	Ostrich	3	13.6	Parakeet	1	4.5
	Eagle	2	9.1	Falcon	1	4.5
	Peacock	2	9.1	Stork	1	4.5
	Emu	2	9.1	Canary	1	4.5
	Budgie	2	9.1	Pigeon	1	4.5
	Chicken	1	4.5			

Nick thought that the *bird* he saw at the zoo was very *colourful*.

FIRST	Parrot	6	50	Parakeet	1	8.3
	Cockatoo	2	16.7	Bird of Paradise	1	8.3
	Peacock	2	16.7			
SECOND	Parrot	2	18.2	Cock	1	9.1
	Puffin	1	9.1	Peacock	1	9.1
	Quetzal	1	9.1	Canary	1	9.1
	Flamingo	1	9.1	Humming Bird	1	9.1
	King Penguin	1	9.1	Budgie	1	9.1

Nick thought that the *colourful bird* he saw at the zoo was very tame.

FIRST	Parrot	8	50	Peacock	2	12.5
	Flamingo	5	31.2	Pelican	1	6.2
SECOND	Parrot	4	26.7	Toucan	1	6.7
	Peacock	3	20	Puffin	1	6.7
	Cockatoo	2	12.5	Humming Bird	1	6.7
	Flamingo	2	12.5	Budgie	1	6.7

The retriever brought back the *bird* that the hunter had seen.

FIRST	Pheasant	6	46.2	Grouse	1	7.7
	Pigeon	5	38.5	Duck	1	7.7
SECOND	Pheasant	3	23.1	Partridge	1	7.7
	Duck	3	23.1	Hawk	1	7.7
	Pigeon	2	15.4	Goose	1	7.7

	Grouse	2	15.4			

The retriever brought back the *bird* that the hunter had seen scavenging.

FIRST	Pheasant	4	33.3	Vulture	2	16.7
	Partridge	2	16.7	Jackdaw	1	11.1
	Grouse	2	16.7	Chaffinch	1	11.1
SECOND	Pheasant	3	25	Crow	1	11.1
	Grouse	3	25	Quail	1	11.1
	Rook	2	16.7	Capercaillie	1	11.1
	Duck	1	11.1			

The retriever brought back the *scavenger bird* that the hunter had seen.

FIRST	Vulture	4	44.4	Crow	1	11.1
	Falcon	1	11.1	Magpie	1	11.1
	Seagull	1	11.1	Hawk	1	11.1
SECOND	Vulture	2	22.2	Seagull	1	11.1
	Gull	1	11.1	Crow	1	11.1
	Starling	1	11.1	Eagle	1	11.1
	Raven	1	11.1			

Although it was unusual the *vegetable* was served with the meal.

FIRST	Courgette	3	17.6	Artichoke	1	5.9
	Aubergine	1	5.9	Tomato	1	5.9
	Sweet Potato	1	5.9	Beansprouts	1	5.9
	Celeriac	1	5.9	Silverbeet	1	5.9
	Lentil	1	5.9	Kumara	1	5.9
	Swede	1	5.9	Broccoli	1	5.9
	Asparagus	1	5.9	Choko	1	5.9
SECOND	Aubergine	2	11.1	Pumpkin	1	5.5
	Artichoke	2	11.1	Fennel	1	5.5
	Broccoli	2	11.1	Leek	1	5.5
	Asparagus	2	11.1	Pepper	1	5.5
	Parsnip	1	5.5	Mango	1	5.5
	Toro	1	5.5	Mung Beans	1	5.5
	Soya	1	5.5	Tomato	1	5.5

Although it was *pale*, the *vegetable* was served with the meal.

FIRST	Parsnip	4	25	Marrow	1	5.5
	Potato	4	25	Red Pepper	1	5.5
	Turnip	3	18.8	Cabbage	1	5.5
	Cauliflower	2	11.1			
SECOND	Swede	6	37.5	Cauliflower	1	5.5
	Turnip	3	18.8	Parsnip	1	5.5
	Cabbage	2	11.1	Tomato	1	5.5
	Potato	1	5.5	Spinach	1	5.5

Although it was unusual the *pale vegetable* was served with the meal.

FIRST	Swede	4	19.0	Potato	1	4.8
	Turnip	3	14.3	Cauliflower	1	4.8
	Cabbage	2	9.5	Yam	1	4.8
	Radish	2	9.5	Parsnip	1	4.8
	Aubergine	2	9.5	Cucumber	1	4.8
	Celery	1	4.8	Lettuce	1	4.8
	Mango	1	4.8			
SECOND	Cabbage	3	15	Kohlrabi	1	5
	Turnip	3	15	Avocado	1	5
	Parsnip	3	15	Yam	1	5
	Potato	2	10	Squash	1	5
	Shallot	1	5	Leek	1	5
	Summer Cabbage	1	5	Courgette	1	5

Because the *vegetable* was used, the meal went further.

FIRST	Potato	10	76.9	Turnip	1	7.7
	Rice	1	7.7	Lettuce	1	7.7
SECOND	Rice	4	33.3	Greens	2	16.7
	Cabbage	2	16.7	Lettuce	1	8.3
	Potato	2	16.7	Pasta	1	8.3

Because the *vegetable* was *cheap* the meal went further.

FIRST	Potato	13	76.5	Carrot	1	5.9
	Cabbage	1	5.9	Turnip	1	5.9
	Rice	1	5.9			
SECOND	Turnip	5	29.4	Rice	1	5.9
	Carrot	3	17.6	Peas	1	5.9
	Potato	3	17.6	Cabbage	1	5.9
	Lettuce	1	5.9	Beans	1	5.9
	Cucumber	1	5.9			

Because the *cheap vegetable* was used the meal went further.

FIRST	Potato	17	68	Carrot	1	4
	Turnip	3	12	Beans	1	4
	Rice	2	8	Cabbage	1	4
SECOND	Cabbage	9	36	Baked Beans	1	4
	Carrot	5	20	Beans	1	4
	Potato	4	16	Turnip	1	4
	Rice	2	8	Lentils	1	4
	Soya Bean	1	4			

The *tool* the fireman used to break into the locked room was a last resort.

FIRST	Axe	16	64	Hammer	1	4
	Hatchet	3	12	Iron Bar	1	4
	Crowbar	2	8	Hand	1	4
	Gun	1	4			
SECOND	Crowbar	8	32	Chainsaw	1	4
	Hammer	5	20	Pick	1	4
	Axe	3	12	Metal Rod	1	4
	Brick	2	8	Mallet	1	4
	Hose	1	4	Hatchet	1	4
	Spade	1	4			

The *tool* the fireman used to break into the locked room was *heavy*.

FIRST	Axe	10	58.8	Hammer	1	5.9
	Pick Axe	2	11.8	Crow Bar	1	5.9
	Sledge Hammer	2	11.8	Chainsaw	1	5.9
SECOND	Sledgehammer	4	26.7	Iron Rod	1	6.7
	Crowbar	3	20	Battering Ram	1	6.7
	Hammer	2	13.3	Fire Hose	1	6.7
	Axe	1	6.7	Wrench	1	6.7

The *heavy tool* the fireman used to break into the locked room was a last resort.

FIRST	Axe	7	53.8	Chair	1	7.7
	Crowbar	2	15.4	Truncheon	1	7.7
	Hammer	1	7.7	Technical Tool	1	7.7
SECOND	Axe	5	38.5	Crowbar	2	15.4
	Sledge Hammer	2	15.4	Parking Meter	1	7.7
	Hammer	2	15.4	Blunt Object	1	7.7

Keith was not allowed to participate because his parents thought the *sport* was unhealthy.

FIRST	Rugby	6	33.3	Swimming	1	6.2
	Jogging	2	12.5	Motorcycling	1	6.2
	Boxing	2	12.5	Competitive Activity	1	6.2
	Mountaineering	1	6.2	Cockfighting	1	6.2
	Chess	1	6.2			
SECOND	Football	4	25	Mountaineering	1	6.2
	Rugby	2	12.5	Squash	1	6.2
	Games	1	6.2	Racing Driving	1	6.2
	Swimming	1	6.2	Darts	1	6.2
	Wrestling	1	6.2	Karate	1	6.2
	Cricket	1	6.2	Badger Baiting	1	6.2

Keith was not allowed to participate because his parents thought the *sport* was cruel.

FIRST	Fox Hunting	9	40.9	Grouse Shooting	1	4.8
	Hunting	4	19.0	Snipes	1	4.8
	Deer Stalking	2	9.5	Rugby	1	4.8
	Boxing	2	9.5	Shooting	1	4.8
SECOND	Fishing	3	13.6	Otter Culling	1	4.5
	Cockfighting	3	13.6	Shooting	1	4.5
	Fox Hunting	2	9.1	Gang Fights	1	4.5
	Deer Hunting	2	9.1	Rugby	1	4.5
	Bull Fighting	2	9.1	Polo	1	4.5
	Hare Baiting	1	4.5	Horse Racing	1	4.5
	Mouse Dissection	1	4.5	Beagling	1	4.5
	Badger Baiting	1	4.5			

Keith was not allowed to participate because his parents thought the *cruel sport* was unhealthy.

FIRST	Foxhunting	6	33.3	Bullfighting	1	5.5
	Rugby	3	16.7	Fishing	1	5.5
	Hunting	2	11.1	Duck Shooting	1	5.5
	Boxing	2	11.1	Grid Iron	1	5.5
	Football	1	5.5			
SECOND	Shooting	4	22.2	Football	1	5.5
	Wrestling	2	11.1	Cockfighting	1	5.5

	Bull Fighting	2	11.1	Rabbiting	1	5.5
	Hare Coursing	2	11.1	Rugby	1	5.5
	Ice Hockey	1	5.5	Show Jumping	1	5.5
	Foxhunting	1	5.5	Polo	1	5.5

He won the *fruit* in a competition.

FIRST	Coconut	6	35.3	Basket of Fruit	1	5.9
	Pineapple	4	23.5	Breadfruit	1	5.9
	Melon	2	11.8	Apple	1	5.9
	Pomegranate	1	5.9	Goods	1	5.9
SECOND	Pineapple	6	37.5	Guava	1	6.2
	Melon	3	18.8	Strawberries	1	6.2
	Banana	2	12.5	Grapes	1	6.2
	Pomegranate	1	6.2	Mango	1	6.2

He won the *fruit* in a competition at the *fairground*.

FIRST	Coconut	17	70.8	Melon	2	8.3
	Apple	3	12.5	Pineapple	2	8.3
SECOND	Pineapple	7	29.2	Toffee Apple	1	4.2
	Coconut	4	16.6	Nuts	1	4.2
	Orange	4	16.6	Pomegranate	1	4.2
	Apple	2	8.3	Banana	1	4.2
	Melon	1	4.2	Passion Fruit	1	4.2
	Grapes	1	4.2			

He won the *fairground fruit* in a competition.

FIRST	Coconut	7	50	Pineapple	1	7.1
	Apple	2	14.3	Banana	1	7.1
	Toffee Apple	2	14.3	Strawberries	1	7.1
SECOND	Orange	3	27.3	Toffee Pear	1	9.1
	Apple	2	18.2	Melon	1	9.1
	Pineapple	2	18.2	Cherries	1	9.1
	Coconut	1	9.1			

Mark got out his *instrument* to play at the lively parade.

FIRST	Trumpet	4	28.6	Flute	1	7.1
	Trombone	2	14.3	Piccolo	1	7.1

	Guitar	1	7.1	Tuba	1	7.1
	Drum	1	7.1	Bugle	1	7.1
	Saxophone	1	7.1	Harmonica	1	7.1
SECOND	Trumpet	5	35.7	Bassoon	1	7.1
	Trombone	2	14.3	Piccolo	1	7.1
	Tuba	2	14.3	Harmonica	1	7.1
	Drum	1	7.1	Saxophone	1	7.1

Mark got out his *instrument* to play at the *Scottish* parade.

FIRST	Bagpipes	14	82.4	Trumpet	1	5.9
	Tin Whistle	1	5.9	Drum	1	5.9
SECOND	Drum	8	50	Tuba	1	6.2
	Fiddle	2	12.5	Bugle	1	6.2
	Bagpipes	2	12.5	Flute	1	6.2
	Mouth Organ	1	6.2			

Mark got out his *Scottish instrument* to play at the lively parade.

FIRST	Bagpipes	23	95.8	Accordion	1	4.2
SECOND	Accordion	7	30.4	Bagpipes	1	4.3
	Fiddle	5	21.7	Whistle	1	4.3
	Drum	3	13.0	Flute	1	4.3
	Clarsach	2	8.7	Fife	1	4.3
	Chanter	1	4.3	Spirtle	1	4.3

APPENDIX 1 Continued

TABLE A3: FREQUENCY LISTS OF EXAMPLES FOR THREE FORMS OF TEN MIXED SENTENCES.

Although it was pleasant he had never tasted the *fruit* before in his life.

	EXEMPLAR	FREQ	%	EXEMPLAR	FREQ	%
FIRST	Kiwi Fruit	6	50	Breadfruit	1	8.3
	Passion Fruit	4	33.3	Guava	1	8.3
SECOND	Kiwi Fruit	4	33.3	Pomegranate	1	8.3
	Mango	2	16.7	Papaya	1	8.3
	Passion Fruit	1	8.3	Raspberry	1	8.3
	Melon	1	8.3	Cumquat	1	8.3

Although it was *exotic* he had never tasted the *fruit* before in his life.

	EXEMPLAR	FREQ	%	EXEMPLAR	FREQ	%
FIRST	Mango	3	33.3	Loquat	1	11.1
	Pawpaw	1	11.1	Cumquat	1	11.1
	Avocado	1	11.1	Lychee	1	11.1
	Pomegranate	1	11.1			
SECOND	Passion Fruit	2	22.2	Lychee	1	11.1
	Pawpaw	1	11.1	Pineapple	1	11.1
	Avocado	1	11.1	Carambola	1	11.1
	Mango	1	11.1	Coconut	1	11.1

Although it was pleasant he had never tasted the *exotic fruit* before in his life.

	EXEMPLAR	FREQ	%	EXEMPLAR	FREQ	%
FIRST	Mango	5	38.5	Passion Fruit	3	23.1
	Kiwi Fruit	4	30.8	Melon	1	7.7
SECOND	Passion Fruit	4	30.8	Pawpaw	1	7.7
	Kiwi Fruit	3	23.1	Melon	1	7.7
	Mango	2	15.4	Pineapple	1	7.7
	Pomegranate	1	7.7			

Bob handled the *bird* with care.

FIRST	Sparrow	4	16.7	Kestrel	2	8.3
	Canary	3	12.5	Raven	1	4.2
	Dove	3	12.5	Chicken	1	4.2
	Seagull	3	12.5	Blackbird	1	4.2
	Budgie	2	8.3	Penguin	1	4.2
	Parrot	2	8.3	Robin	1	4.2
SECOND	Parrot	3	14.3	Starling	1	4.8
	Dove	3	14.3	Duck	1	4.8
	Budgie	2	9.5	Mynah	1	4.8
	Pigeon	2	9.5	Eagle	1	4.8
	Sparrow	2	9.5	Hen	1	4.8
	Falcon	2	9.5	Canary	1	4.8

It was *large* so Bob handled the *bird* with care.

FIRST	Eagle	7	31.8	Owl	1	4.5
	Ostrich	2	9.1	Emu	1	4.5
	Seagull	2	9.1	Kestrel	1	4.5
	Swan	2	9.1	Chicken	1	4.5
	Vulture	2	9.1	Hawk	1	4.5
	Penguin	1	4.5	Pigeon	1	4.5
SECOND	Eagle	5	22.7	Hawk	1	4.5
	Seagull	3	13.6	Parrot	1	4.5
	Albatross	3	13.6	Pelican	1	4.5
	Flamingo	2	9.1	Pigeon	1	4.5
	Osprey	2	9.1	Crow	1	4.5
	Kestrel	1	4.5	Duck	1	4.5

Bob handled the *large bird* with care.

FIRST	Eagle	6	26.1	Seagull	1	4.3
	Ostrich	3	13.0	Crow	1	4.3
	Parrot	3	13.0	Sparrow	1	4.3
	Swan	2	8.7	Vulture	1	4.3
	Emu	2	8.7	Owl	1	4.3
	Falcon	1	4.3	Pelican	1	4.3
SECOND	Osprey	3	13.0	Owl	1	4.3
	Ostrich	3	13.0	Dove	1	4.3
	Seagull	2	8.7	Albatross	1	4.3
	Crow	2	8.7	Flamingo	1	4.3
	Parrot	2	8.7	Falcon	1	4.3
	Swan	2	8.7	Pheasant	1	4.3
	Emu	1	4.3	Chaffinch	1	4.3
	Duck	1	4.3			

Carol examined it and decided that the *instrument* could be fine.

FIRST	Flute	4	18.2	Oboe	1	4.5
	Piano	4	18.2	Harp	1	4.5
	Scalpel	4	18.2	Scissors	1	4.5
	Violin	3	13.6	Compass	1	4.5
	Trumpet	1	4.5	Pick Axe	1	4.5
	Syringe	1	4.5			
SECOND	Violin	3	14.3	Saxophone	1	4.8
	Guitar	2	9.5	Trombone	1	4.8
	Scalpel	2	9.5	Clock	1	4.8
	Drum	2	9.5	Kitchen Knife	1	4.8
	Recorder	1	4.8	Tuning Fork	1	4.8
	Needle	1	4.8	Screwdriver	1	4.8
	Clarinet	1	4.8	Organ	1	4.8
	Tuba	1	4.8	Oboe	1	4.8

Carol examined it *unstrung* and decided that the *instrument* could be fine.

FIRST	Violin	10	43.5	Viola	2	8.7
	Guitar	6	26.1	Banjo	1	4.3
	Cello	3	13.0	Microscope	1	4.3
SECOND	Violin	7	30.4	Piano	1	4.3
	Cello	7	30.4	Microwave	1	4.3
	Guitar	3	13.0	Double Bass	1	4.3
	Harp	2	8.7	Banjo	1	4.3

Carol examined it and decided that the *unstrung instrument* could be fine.

FIRST	Guitar	9	40.9	Cello	1	4.5
	Violin	7	31.8	French Horn	1	4.5
	Harp	2	9.1	Fiddle	1	4.5
	Piano	1	4.5			
SECOND	Violin	7	31.8	Viola	2	9.1
	Guitar	4	18.2	Double Bass	1	4.5
	Cello	4	18.2	Tuba	1	4.5
	Harp	3	13.6			

Charles carefully cleaned the *instrument* he had found.

FIRST	Flute	5	23.8	Hatchet	1	4.8
	Trumpet	3	14.3	Watch	1	4.8
	Gun	3	14.3	Screwdriver	1	4.8
	Tin Whistle	2	9.5	Telescope	1	4.8
	Sword	1	4.8	Fyfe	1	4.8
	Scalpel	1	4.8	Euphonium	1	4.8

SECOND	Recorder	3	14.3	Gun	1	4.8
	Flute	2	9.5	Icepick	1	4.8
	Knife	2	9.5	Blender	1	4.8
	Microscope	2	9.5	Compass	1	4.8
	Clarinet	2	9.5	Scales	1	4.8
	Guitar	2	9.5	Lute	1	4.8
	Trumpet	1	4.8	Trombone	1	4.8

Charles carefully *disassembled* and cleaned the *instrument* he had found.

FIRST	Trumpet	5	22.7	Chisel	1	4.5
	Flute	4	18.2	Chainsaw	1	4.5
	Saxophone	2	9.1	Microscope	1	4.5
	Clarinet	2	9.1	Gun	1	4.5
	Oboe	1	4.5	Shotgun	1	4.5
	Pipe	1	4.5	Clock	1	4.5
	Drum	1	4.5			
SECOND	Trumpet	5	22.7	Hornet	1	4.5
	Clarinet	3	13.6	Watch	1	4.5
	Trombone	3	13.6	Plane	1	4.5
	Oboe	2	9.1	Camera	1	4.5
	Flute	2	9.1	Gramophone	1	4.5
	Saxophone	2	9.1			

Charles carefully cleaned the *disassembled instrument* he had found.

FIRST	Flute	5	21.7	Car	1	4.3
	Clarinet	4	17.4	Watch	1	4.3
	Clock	2	8.7	Violin	1	4.3
	Gun	2	8.7	Trombone	1	4.3
	Bicycle	1	4.3	Theodolite	1	4.3
	Stethoscope	1	4.3	Saxophone	1	4.3
	Piano	1	4.3	Pen	1	4.3
SECOND	Clarinet	6	26.1	Bicycle	1	4.3
	Trombone	2	8.7	Toy train	1	4.3
	Microscope	2	8.7	Violin	1	4.3
	Flute	2	8.7	Bagpipes	1	4.3
	French Horn	1	4.3	Saxophone	1	4.3
	Recorder	1	4.3	Guitar	1	4.3
	Telescope	1	4.3	Motorbike	1	4.3
	Watch	1	4.3			

Although he was still by himself, the *animal* was entertaining.

FIRST	Monkey	6	28.6	Sealion	1	4.8

	Dog	4	19.0	Orangutan	1	4.8
	Kitten	2	9.5	Zebra	1	4.8
	Elephant	2	9.5	Seal	1	4.8
	Lamb	1	4.8	Mouse	1	4.8
	Lion	1	4.8			
SECOND	Dog	3	13.6	Penguin	1	4.5
	Monkey	2	9.1	Parrot	1	4.5
	Kitten	2	9.1	Chimpanzee	1	4.5
	Cat	2	9.1	Seal	1	4.5
	Elephant	2	9.1	Polar Bear	1	4.5
	Budgie	1	4.5	Dolphin	1	4.5
	Gorilla	1	4.5	Child	1	4.5
	Horse	1	4.5	Puppy	1	4.5

Although he was still *dangling* by himself, the *animal* was entertaining.

FIRST	Monkey	13	56.5	Bear	1	4.3
	Chimpanzee	2	8.7	Mouse	1	4.3
	Puppy	1	4.3	Seal	1	4.3
	Bat	1	4.3	Gibbon	1	4.3
	Lemur	1	4.3	Orangutan	1	4.3
SECOND	Sloth	4	17.4	Koala Bear	1	4.3
	Gibbon	3	13.0	Marmoset	1	4.3
	Gorilla	2	8.7	Lion	1	4.3
	Snake	2	8.7	Cat	1	4.3
	Orangutan	1	4.3	Chimpanzee	1	4.3
	Ape	1	4.3	Dog	1	4.3
	Hummingbird	1	4.3	Bear	1	4.3
	Bat	1	4.3			

Although he was still by himself, the *dangling animal* was entertaining.

FIRST	Monkey	9	40.9	Chinchilla	1	4.5
	Chimpanzee	3	13.6	Gibbon	1	4.5
	Bat	3	13.6	Rabbit	1	4.5
	Baboon	2	9.1	Hamster	1	4.5
	Sloth	1	4.5			
SECOND	Chimpanzee	3	13.6	Orangutan	1	4.5
	Snake	3	13.6	Parrot	1	4.5
	Spider	3	13.6	Penguin	1	4.5
	Koala Bear	2	9.1	Bat	1	4.5
	Sloth	1	4.5	Monkey	1	4.5
	Gibbon	1	4.5	Possum	1	4.5
	Clown	1	4.5	Dog	1	4.5
	Bushbaby	1	4.5			

Capturing everyone's attention the *animal* was the main attraction.

FIRST	Lion	8	36.4	Arab Horse	1	4.5
	Monkey	3	13.6	Zebra	1	4.5
	Dog	3	13.6	Penguin	1	4.5
	Chimpanzee	2	9.1	Seal	1	4.5
	Panther	1	4.5	Elephant	1	4.5
SECOND	Monkey	3	13.6	Lion	1	4.5
	Elephant	3	13.6	Ostrich	1	4.5
	Dolphin	2	9.1	Cheetah	1	4.5
	Seal	2	9.1	Gorilla	1	4.5
	Dog	2	9.1	Polar Bear	1	4.5
	Tiger	2	9.1	Dolphin	1	4.5
	Baboon	1	4.5	Bull	1	4.5

Making a *noise* and capturing everyone's attention the *animal* was the main attraction.

FIRST	Lion	7	33.3	Sea Lion	1	4.8
	Monkey	4	19.0	Dog	1	4.8
	Elephant	2	9.5	Tiger	1	4.8
	Chimpanzee	2	9.5	Cat	1	4.8
	Parrot	2	9.5			
SECOND	Lion	5	23.8	Tiger	2	9.5
	Monkey	4	19.0	Baby	1	4.8
	Elephant	4	19.0	Hyena	1	4.8
	Seal	3	14.3	Bird	1	4.8

Capturing everyone's attention the *noisy animal* was the main attraction.

FIRST	Lion	4	17.4	Chimpanzee	1	4.3
	Monkey	4	17.4	Hyena	1	4.3
	Dog	4	17.4	Baboon	1	4.3
	Elephant	2	8.7	Puppy	1	4.3
	Sealion	1	4.3	Gorilla	1	4.3
	Cheetah	1	4.3	Walrus	1	4.3
	Bear	1	4.3			
SECOND	Lion	5	23.8	Cat	1	4.8
	Tiger	3	14.3	Seal	1	4.8
	Elephant	1	4.8	Giraffe	1	4.8
	Cow	1	4.8	Baby	1	4.8
	Gorilla	1	4.8	Hyena	1	4.8
	Seagull	1	4.8	Monkey	1	4.8
	Dog	1	4.8	Orangutan	1	4.8
	Chimpanzee	1	4.8			

"He's over there" indicated the keeper, pointing at the *animal* in the field.

FIRST	Horse	4	17.4	Highland Cow	1	4.3
	Cow	4	17.4	Tiger	1	4.3
	Lion	3	13.0	Stag	1	4.3
	Bull	2	8.7	Zebra	1	4.3
	Camel	2	8.7	Donkey	1	4.3
	Deer	1	4.3	Dog	1	4.3
	Sheep	1	4.3			
SECOND	Horse	5	21.7	Donkey	1	4.3
	Cow	4	17.4	Dog	1	4.3
	Zebra	3	13.0	Bull	1	4.3
	Giraffe	2	8.7	Goat	1	4.3
	Cat	2	8.7	Sheep	1	4.3
	Tiger	1	4.3	Deer	1	4.3

"He's *hiding* over there" indicated the keeper pointing at the *animal* in the field.

FIRST	Fox	4	18.2	Deer	1	4.5
	Dog	2	9.1	Tiger	1	4.5
	Lion	2	9.1	Buffalo	1	4.5
	Goat	2	9.1	Penguin	1	4.5
	Horse	2	9.1	Leopard	1	4.5
	Cow	2	9.1	Pheasant	1	4.5
	Giraffe	1	4.5	Bull	1	4.5
SECOND	Dog	3	13.6	Bear	1	4.5
	Rabbit	2	9.1	Cow	1	4.5
	Sheep	2	9.1	Elephant	1	4.5
	Horse	2	9.1	Goat	1	4.5
	Ewe	1	4.5	Wolf	1	4.5
	Badger	1	4.5	Polar Bear	1	4.5
	Tiger	1	4.5	Hedgehog	1	4.5
	Rat	1	4.5	Fox	1	4.5
	Pig	1	4.5			

"He's over there" indicated the keeper, pointing at the *hiding animal* in the field.

FIRST	Rabbit	6	27.3	Wolf	1	4.5
	Fox	5	22.7	Koala Bear	1	4.5
	Dog	2	9.1	Bull	1	4.5
	Tiger	2	9.1	Lion	1	4.5
	Deer	1	4.5	Horse	1	4.5
	Badger	1	4.5			
SECOND	Rabbit	5	22.7	Elephant	1	4.5
	Racoon	1	4.5	Lion	1	4.5
	Hyena	1	4.5	Ram	1	4.5
	Sheep	1	4.5	Rat	1	4.5
	Tiger	1	4.5	Hedgehog	1	4.5
	Antelope	1	4.5	Horse	1	4.5
	Goat	1	4.5	Deer	1	4.5

	Fox	1	4.5	Badger	1	4.5
	Ferret	1	4.5	Pheasant	1	4.5

Seldom as it was, Sam played the *sport* well.

FIRST	Football	8	34.8	Badminton	1	4.3
	Tennis	4	17.4	Golf	1	4.3
	Cricket	3	13.0	Fencing	1	4.3
	Squash	2	8.7	Croquet	1	4.3
	Rugby	2	8.7			
SECOND	Cricket	6	27.3	Volleyball	1	4.5
	Tennis	5	22.7	Ice Hockey	1	4.5
	Squash	2	9.1	Bowls	1	4.5
	Rugby	2	9.1	Snooker	1	4.5
	Badminton	2	9.1	Football	1	4.5

Difficult as it was, Sam played the *sport* well.

FIRST	Squash	4	19.0	Badminton	1	4.8
	Football	3	14.3	Lacrosse	1	4.8
	Tennis	3	14.3	Volleyball	1	4.8
	Polo	2	9.5	Basketball	1	4.8
	Golf	2	9.5	Canoeing	1	4.8
	Fives	1	4.8	Rugby	1	4.8
SECOND	Rugby	3	15	Volleyball	1	5
	Cricket	3	15	Basketball	1	5
	Table Tennis	2	10	Squash	1	5
	Hockey	2	10	Judo	1	5
	Tennis	2	10	Badminton	1	5
	Golf	2	10	Football	1	5

Seldom as it was, Sam played the *difficult sport* well.

FIRST	Polo	5	23.8	Cricket	2	9.5
	Squash	4	19.0	Volleyball	1	4.8
	Rugby	3	14.3	Snooker	1	4.8
	Lacrosse	2	9.5	Croquet	1	4.8
	Tennis	2	9.5			
SECOND	Hockey	3	14.3	Billiards	1	4.8
	Squash	3	14.3	Lacrosse	1	4.8
	Rugby	2	9.5	Badminton	1	4.8
	Tennis	2	9.5	Croquet	1	4.8
	Snooker	1	4.8	Volleyball	1	4.8
	Cricket	1	4.8	Football	1	4.8
	Fencing	1	4.8	Chess	1	4.8

The season encouraged brisk sales of *clothing*.

FIRST	Jumper	3	18.8	T-shirt	1	4.2
	Coat	3	18.8	Swimwear	1	4.2
	Shoes	2	12.5	Hat	1	4.2
	Shorts	2	12.5	Pullover	1	4.2
	Summer Clothes	1	4.2	Blouse	1	4.2
SECOND	Swimwear	3	18.7	Raincoat	1	6.2
	Coat	2	12.5	Hat	1	6.2
	Boots	2	12.5	Sun Hat	1	6.2
	Gloves	2	12.5	Shirt	1	6.2
	Sandals	1	6.2	Trousers	1	6.2
	Suit	1	6.2			

The *summer* season encouraged brisk sales of *clothing*.

FIRST	T-shirt	5	22.7	Cotton Dress	2	9.1
	Bikini	4	18.2	Skirt	1	4.5
	Swimming Costume	3	13.6	Jacket	1	4.5
	Shorts	3	13.6	Umbrella	1	4.5
	Trunks	2	9.1			
SECOND	T-shirt	6	27.3	Overcoat	1	4.5
	Swimwear	4	18.2	Jean Shorts	1	4.5
	Shorts	4	18.2	Shirt	1	4.5
	Shoes	2	9.1	Skirt	1	4.5
	Underwear	1	4.5	Wellies	1	4.5

The season encouraged brisk sales of *summer clothing*.

FIRST	Shorts	6	30	Shoes	1	5
	T-shirt	5	25	Skirt	1	5
	Shirt	3	15	Hat	1	5
	Dress	2	10	Swimsuit	1	5
SECOND	T-shirt	5	25	Hat	1	5
	Bikini	4	20	Blouse	1	5
	Swimming Trunks	2	10	Shirt	1	5
	Dress	2	10	Skirt	1	5
	Shorts	2	10	Sun Top	1	5

The *vegetable* tasted very strange when eaten raw.

FIRST	Potato	9	40.9	Marrow	1	4.5
	Cauliflower	3	13.6	Parsnip	1	4.5
	Onion	2	9.1	Chilli (pepper)	1	4.5
	Courgette	2	9.1	Garlic	1	4.5
	Cabbage	2	9.1			
SECOND	Potato	5	23.8	Onion	1	4.8
	Cauliflower	3	14.3	Celery	1	4.8
	Carrot	2	9.5	Courgette	1	4.8
	Beetroot	1	4.8	Asparagus	1	4.8
	Peas	1	4.8	Pumpkin	1	4.8
	Turnip	1	4.8	Runner Beans	1	4.8
	Leek	1	4.8	Cabbage	1	4.8

The *vegetable* tasted very strange when eaten raw from the *garden*.

FIRST	Carrot	5	21.7	Brussels Sprout	2	8.7
	Turnip	4	17.4	Spinach	1	4.3
	Potato	3	13.0	Marrow	1	4.3
	Beetroot	3	13.0	Courgette	1	4.3
	Radish	2	8.7	Cabbage	1	4.3
SECOND	Potato	5	21.7	Cabbage	1	4.3
	Parsnip	3	13.0	Beans	1	4.3
	Carrot	3	13.0	Celery	1	4.3
	Beetroot	2	8.7	Rhubarb	1	4.3
	Turnip	2	8.7	Lettuce	1	4.3
	Broccoli	1	4.3	Brussels Sprout	1	4.3
	Peas	1	4.3			

The *garden vegetable* tasted very strange when eaten raw.

FIRST	Potato	5	22.7	Marrow	1	4.5
	Carrot	3	13.6	Green Pepper	1	4.5
	Peas	2	9.1	Parsnip	1	4.5
	Cabbage	2	9.1	Spinach	1	4.5
	Onion	1	4.5	Broad Beans	1	4.5
	Turnip	1	4.5	Leek	1	4.5
	Rhubarb	1	4.5	Courgette	1	4.5
SECOND	Turnip	5	22.7	Carrot	1	4.5
	Potato	5	22.7	Brussels Sprout	1	4.5
	Spinach	2	9.1	Broccoli	1	4.5
	Lettuce	1	4.5	Cauliflower	1	4.5
	Asparagus	1	4.5	Tomato	1	4.5
	Parsnip	1	4.5	Pumpkin	1	4.5
	Beans	1	4.5			

Although it was the nearest one to hand the *weapon* proved effective.

FIRST	Club	1	11.1	Candlestick	1	11.1
	Axe	1	11.1	Gun	1	11.1
	Cannon	1	11.1	Knife	1	11.1
	Kitchen Knife	1	11.1	Vase	1	11.1
	Plate	1	11.1			
SECOND	Iron Bar	1	11.1	Hammer	1	11.1
	Cleaver	1	11.1	Gun	1	11.1
	Stone	1	11.1	Screwdriver	1	11.1
	Chair	1	11.1	Shoe	1	11.1

Although it was the nearest one to hand the *weapon* proved *deadly*.

FIRST	Knife	4	30.8	Poker	1	7.7
	Gun	2	15.4	Paper Weight	1	7.7
	Screwdriver	1	7.7	Stone	1	7.7
	Letter Opener	1	7.7	Bottle	1	7.7
	Fork	1	7.7			
SECOND	Wooden Beam	1	7.7	Crossbow	1	7.7
	Poison Dart	1	7.7	Gun	1	7.7
	Spade	1	7.7	Penknife	1	7.7
	Poker	1	7.7	Vase	1	7.7
	Knife	1	7.7	Paper Knife	1	7.7
	Bottle	1	7.7	Beer Can	1	7.7

Although it was the nearest one to hand the *deadly weapon* proved effective.

FIRST	Knife	5	45.4	Scissors	1	9.1
	Bomb	1	9.1	Tank	1	9.1
	Gun	1	9.1	Dagger	1	9.1
	Letter Opener	1	9.1			
SECOND	Gun	2	18.2	Machine Gun	1	9.1
	Knife	1	9.1	Samurai Sword	1	9.1
	Garden Shears	1	9.1	Cheese Wire	1	9.1
	Dagger	1	9.1	Iron Bar	1	9.1
	Knitting Needle	1	9.1	Lamp	1	9.1

APPENDIX 2

Are Contextual Changes Conceptual Changes?

Introduction

Many authors have reported that context affects which exemplars are considered prototypical of a category, and Experiment 1 served to clarify the wide extent of these contextual alterations. A possible implication of this finding is that concepts themselves alter with context. If concepts do change with context then we would expect contextual prototypes to affect cognitive processes to a similar extent to Rosch's prototypes. That is, if prototypes, as documented by Rosch, had an effect on cognitive decisions and problem solving, these context specific prototypes should have similar cognitive effects. Rosch examined the cognitive effects of intuitive goodness of example distributions experimentally, in order to establish their credibility as important psychological variables. If similar claims are to be made about context specific prototypes, experimental verification of their explanatory power is essential. The question which needs to be addressed experimentally is whether the change in reported prototypicality with context is a result of change at the conceptual level.

Whitney & Kellas (1984) published a paper in which they claimed that there are no contextual changes of prototypicality at the conceptual level. That is to say if, for example, a subject was primed with a sentence suggesting an atypical instantiation of a category name that sentence would prime the category prototype rather than the context dependent prototype. So, if presented with a sentence such as: "The State authorities used their chosen weapon to execute criminals humanely", Whitney & Kellas argue that the prototype "gun" would be primed rather than the atypical, but context-appropriate, "electric chair". It is argued that context effects such as those documented in Experiment 1 are the result of "post-access inferences" (also suggested by Gumenik, 1979): that is inferences about likely referents and meaning made after the concept has been accessed. Whitney & Kellas argue that inferences are not routinely made during the encoding of individual sentences, and cite Dosher & Corbett (Corbett & Dosher, 1978; Dosher & Corbett, 1982) in their defence.

Previous to Dosher & Corbett it had been argued that highly likely implicit referents were routinely inferred during reading (Paris & Lindauer, 1976). For example for sentences such as: "The lawyer cooked dinner on a stove" and "The lawyer cooked dinner", "stove" was an equally good recall cue. Thus, implicit instances serve as equally good recall cues as instances that are mentioned explicitly. Corbett & Dosher (1978) argued that this effect was a result of post-access inference, rather than a result of the implicit instance being encoded when the sentence is first read. They argued instead that implicit instances "may serve as effective cues by virtue of their representation in... abstract schemata". Dosher & Corbett conducted an experiment which included a third form of sentence which incorporated a low probability instrument; e.g. "The lawyer cooked dinner on a campfire". Paris & Lindauer would predict that "campfire" would be encoded in place of "stove" when the sentence was read, and that "stove" would not act as a retrieval cue. Corbett & Dosher found, however, that "stove" acted as a recall cue for the sentence in which a low probability item was encoded as the instance. Thus they concluded that inferences (of this sort) are not routinely made during encoding. It follows from this that changes in typicality due to context may also be a result of post-access inference.

Whitney & Kellas use this rationale to argue that if context effects are present at the encoding stage then a conceptual change must have taken place. Conversely, if context effects only emerge at a later stage of processing they must be the result of post-access inference or, as Barsalou suggests, post-access combination. Whitney & Kellas (1984) designed an experimental paradigm in which individual sentences served as context dependent primes. Thus, context sentences were presented to subjects in order to test whether they primed typical or atypical context-relevant exemplars. Whitney & Kellas utilized a Stroop task (Stroop, 1938) after Warren (1972) and Conrad (1974) who had shown that the greater the association between the word printed in coloured ink and the prime the longer subjects take to name the colour. For the Stroop effect to occur it is unnecessary for the target word to be another colour name, the activation from the prime of an entirely separate word is sufficient to cause interference. The Stroop paradigm has proved a sensitive measure of priming, and was used by Whitney & Kellas to investigate whether context specific prototypes were routinely primed when encoding sentences which intuitively suggested them.

Whitney & Kellas reported that when this test was used context specific prototypes were found to have no effect on cognition. Moreover, prototypes documented without context (i.e. those reported by Rosch and others) were still primed by atypical sentences. This effect could not be explained by saying that subjects were simply not attending sufficiently to the sentence context because subjects were informed that they would be given a sentence recognition test at the end of each block. Thus subjects were motivated to direct their attention to understanding the whole sentence. If reliable, Whitney & Kellas' result is an important one because it strongly supports theories of semantic memory which suggest concepts have a prototypical structure. If context effects are just a result of post-access inference processes, then documenting and understanding prototypes obtained in a 'neutral' context, where context is unspecified, is an important basis on which a theory of meaning can be built. It would imply that we have stable concepts which form the building blocks of cognition, and that all we need discover in addition are the appropriate rules of manipulation (or perhaps of concept conjunction) to fully understand how and why word meanings change in context. A post-access inference explanation thus displaces the role of context, one of the most difficult and challenging factors to models of language and meaning, to the status of a side issue with regard to theoretical questions.

The question of whether or not contextual changes necessitate conceptual changes has been shown to be theoretically important. This experiment is therefore designed as a replication in kind of Whitney & Kellas (1984). Slightly modified conditions are used to ascertain whether the effects previously found are robust enough to generalize to new sentences and new equipment.

METHOD

Design

The design of the experiment was within subjects. The independent variables were sentence type and target type. Sentence type had three levels: sentences suggesting typical category members, atypical category members, and control sentences (which contained a category name which was unrelated to the target word). Target type had two levels: words normally considered typical category members and words normally considered atypical category members. Although target words varied in length and frequency, factors which might have been expected to affect response latency, these differences were controlled for in the experimental design by including control sentences. A prime facilitates response times by providing a warning of the stimulus onset, and a clue about the information to come (Rosch, 1975a). The signal, or warning function of the prime in this case was taken to be the response time of a subject to a target in the control condition (when the target is preceded by an unrelated sentence). This provides information about how factors such as word frequency and word length might affect reaction times. The information function of the prime was taken to be the the response times for the same target word when primed with a related sentence, minus the response time when the target was preceded by a control sentence. Priming increases response latency by setting up a response competition between the expected target and the name of the colour. Thus we would expect the decision time to be longer when the target word was preceded by a related sentence than when it was preceded by a control sentence.

Targets were printed in four colours red, blue, yellow and purple. Colours were chosen to maximize distinctiveness between colours when displayed on the computer screen. As target colour might be expected to affect colour identification latencies due to colour preference, or even fluke associations with sentence stimuli, possible effects were controlled by randomly assigning the four colours to targets. Randomization was such that each colour was presented equally often. A different random order of target colours was used for each subject.

Subjects

Eighteen people volunteered to take part in the experiment. Subjects were students at the University of Edinburgh. None of the subjects had participated in Experiment 1. The sexes were approximately equally represented. All subjects were native English speakers. Subjects reported to have normal, or corrected to normal, vision and normal colour vision.

Stimuli

20 target sentences used in the experiment were taken from the norms of Experiment 1. 10 typical and 10 atypical sentences were chosen. Sentences from the "instrument" category were only selected if the sentence unambiguously suggested a musical instrument. In addition 30 control sentences were constructed in the same way as the sentences used in Experiment 1. The control sentences were thus similar in linguistic form to, but different in semantic meaning from, the experimental sentences. 10 control sentences were used to form a practice run prior to the experiment. The target and control sentences reflected the three linguistic forms used in Experiment 1.

Sentence-word pairs were constructed for typical, atypical and control sentences. Typical sentences were paired with typical target words cited as good exemplars in Experiment 1, making sentence and target highly related. The same sentences were also paired with atypical category members (taken from Uyeda & Mandler, 1980) these pairings were thus poorly related. Similarly, atypical sentences were paired firstly with typical category members (from Uyeda & Mandler, 1980) and forming low-related pairs; and secondly with atypical category members cited as good examples for atypical sentences in Experiment 1, and thus highly related. Sentence-target pairs were divided into two lists such that each sentence appeared only once per list.

Twenty control sentences were paired with the typical and atypical target words from the opposite list so that target words only appeared once per list. The 10 control sentences for the practice run were paired with 10 new nouns, the control sentences were unrelated to the targets in the practice session. List 1 contained: 20 control sentences paired with 10 unrelated typical category instances and 10 unrelated atypical category instances; and 20 priming sentences: 10 typical sentences half paired with related typical targets, half with related atypical targets; and 10 atypical sentences half paired with related typical targets, half with related atypical targets. List 2 was the mirror image of List 1: typical sentences previously paired with typical instances were now paired with atypical instances. 20 control sentences were also repeated, now paired with the alternative unrelated word. Controls and sentences were randomized and presented to the subject alternately (after Beller, 1971; Rosch, 1973, 1975a). The same order of presentation was used for each subject.

Apparatus and Materials

The experiment was conducted on a BBC Master microcomputer with a colour CUB microvitec 452 monitor. A four key response box was connected to the computer. The response keys were 2cm wide and were labelled with four colour names which read from left to right: RED, BLUE, YELLOW and PURPLE. Colour names were stencilled in normatype 22.12 CLN, 3.2mm capital letters on a white background. Subjects used their dominant hand to press the keys. All those who took part in the experiment were right handed.

Four recognition sheets were prepared. Each sheet contained 10 sentences, 5 were priming and control sentences from the experiment and 5 were distractor sentences. Distractor sentences were constructed in the same way as control sentences for the experiment but had never been seen before by the subjects. Sentences were typed in a random order on a sheet of A4 paper.

Procedure

Each subject was tested in one experimental session which lasted approximately 30 minutes. Each session consisted of 10 practice trials and 40 experimental trials. Subjects were tested singly in a laboratory. They were seated at a table 44 cm wide on which the response keys lay. The monitor was placed on a bench behind the table, the midpoint of the monitor screen was 106 cm high, that is approximately eye-level to the seated subject. Subjects were randomly assigned to List 1 or List 2.

The room was darkened to reduce reflection on the monitor screen. The experimenter explained to the subject that the aim of the research was to investigate the effects of one task upon another. Subjects were told their first task was to read, understand and remember the sentences which appeared on the screen; and their second task was to respond as quickly and accurately as possible to the colour of the word displayed after each sentence by pressing the appropriately labelled response key. The subject was told that a recognition sheet would be presented at the end of each block.

The experiment was divided into a practice session and 4 experimental sections each of 10 sentences. After the practice block the decision times and errors for the practice were displayed on the computer screen, and the subjects encouraged to respond more quickly or accurately as appropriate. After each block subjects were given a recognition sheet on which they indicated which sentences they had seen in the previous section. The computer was programmed to stop automatically at the end of each section, and was restarted by the experimenter after the recognition sheet was completed.

Subjects were given as long as necessary to complete the recognition sheets. Sentences were displayed in the centre of the monitor screen for 6s, in normal white print. There was a 500ms delay between sentence erasure and target onset. Targets were displayed in capital letters in the centre of the screen in one of four colours, and remained there until a response was made. There was a 500ms delay before the next sentence was displayed. Colour identification latencies were recorded by the computer.

RESULTS

Sentence Recognition

All subjects performed very well on the recognition task, indicating that close attention was paid to priming sentences. Average error rates were 6.4%. When errors were made they were almost always errors of omission, that is a failure to identify sentences which were presented rather than wrongly selecting new sentences. This suggests that where errors occurred they may have been due to strict selection of sentences rather than complete lack of recall.

Colour Identification Latencies

Average error rates were 3% and did not vary significantly for conditions. Errors were excluded from further analyses. Following Whitney & Kellas (1984) mean response times for each subject were analysed in a 3 X 2 (sentence type X target type) within subjects analysis of variance (ANOVA). Average response times for lists 1 and 2 were 827ms and 839ms; this difference was not significant.

Average response times and error rates for each condition are presented in Table 1. No main effects or group differences were found to be significant.

TARGET TYPE	SENTENCE PRIME		
	Control	Typical	Atypical
Typical	841 (1.1)	852 (6.6)	837 (3.3)
Atypical	815 (4.4)	826 (0)	827 (3.3)

TABLE 1
Average response times (in ms) of subjects to colour identification with different priming and target conditions. (Average % error rates are shown in parenthesis).

Materials Analysis

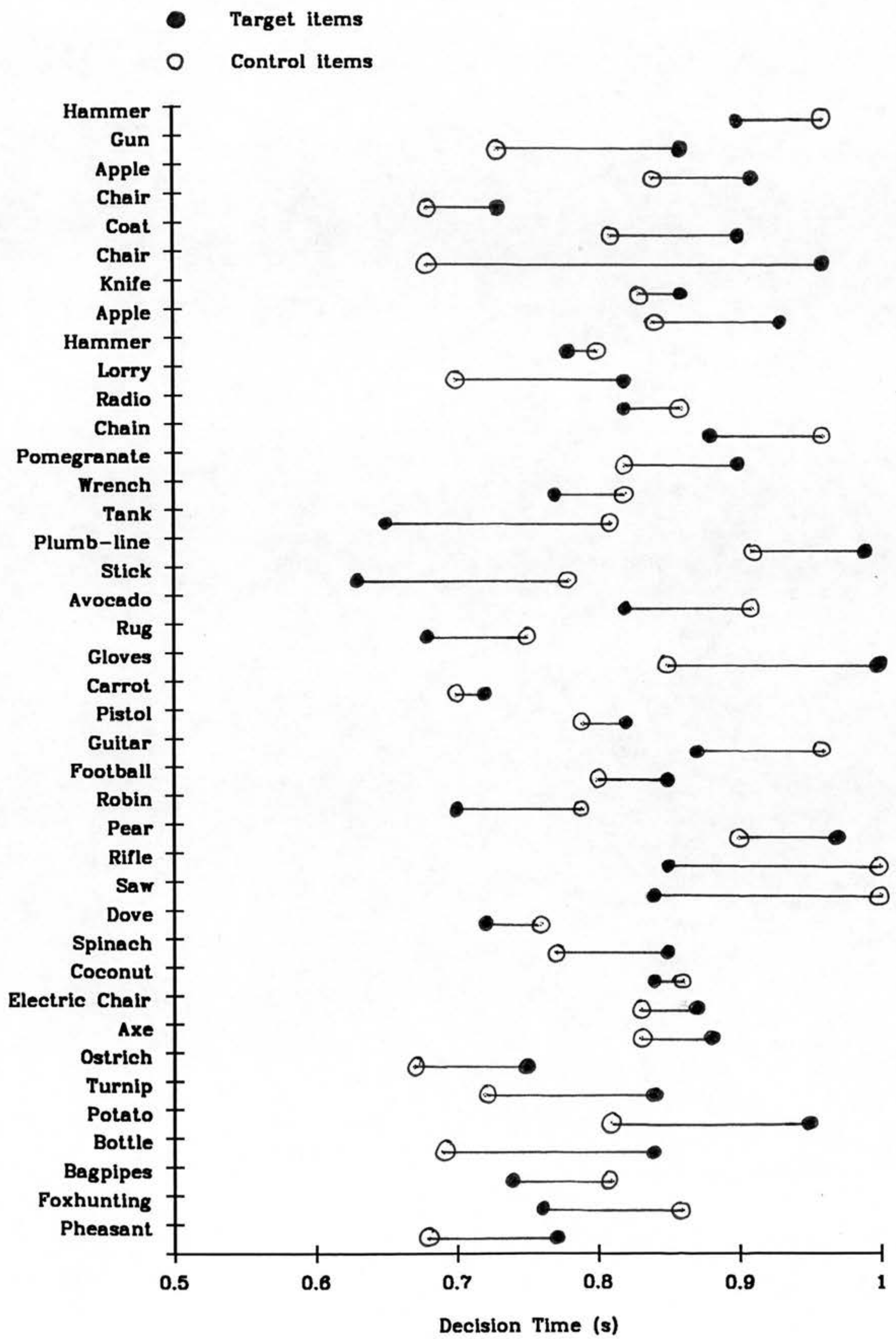
The possibility of carrying out ad hoc analysis of materials was investigated but was found to be inappropriate in this particular case. This was partly because individual words were only repeated across subjects (and then only once). Materials were random across conditions, but order effects could be expected to have a significant influence on the response times for particular items. These factors prohibited a statistical analysis on materials to be performed, but it was felt that some idea of the possible contribution of individual items could be gained from the presentation of mean reaction times for each word, providing the above qualifications are taken seriously.

The average decision time for each word in control (paired with an unrelated sentence) and target (paired with a related sentence) conditions are presented in Graph 1. Note that the graph includes materials from both lists 1 and 2 (a between subjects factor) and, in particular, that subjects received each word either in a control condition or a target condition. Words are ordered in groups so that between group variation can be compared with within group variation, and any main effects clearly demonstrated. An inspection of Graph 1 confirms that there was considerable variation within conditions and no clear pattern between conditions.

There is no indication in Graph 1 that any particular items generated an unusual

GRAPH 1

Response times for items in target (related) and control (unrelated) conditions



reponse from subjects. Examination of how response times to individual words varied in the control and target conditions reveals no effect: 22 items had a faster average response time when in the target condition, while 18 had a faster average response time in the control condition. There is, however, a suggestion that the target type is a significant factor for typical sentences. Eight out of ten typical words had longer decision times when primed by a related typical sentence, while seven out of ten atypical words had longer decision times when primed by an unrelated typical sentence. This pattern was not seen in the data for atypical priming sentences, where target and controls are equally likely to be responded to quickly. Further study is needed to decide whether this pattern has simply occurred by chance.

Discussion

This experiment did not provide a replication of Whitney & Kellas' finding that context sentences primed "neutral" prototypes rather than context dependent prototypes. When these results are compared with those of Whitney & Kellas (1984) it can be seen that the pattern of results obtained (see Table 1) is very similar to the pattern they report (see Table 2). Results for this experiment were on average 104ms faster than found by Whitney & Kellas. This difference can be attributed to task differences, such as the voice onset delay effect incorporated in Whitney & Kellas' naming task. Error rates are similar in both experiments. The one striking difference between these results and those reported by Whitney & Kellas occurs in the control - typical condition. Whitney & Kellas report that this response time is 890ms, which is 20ms faster than when the control was paired with an atypical instance (910ms). Here, however, the control - typical time (841ms) is 26ms slower than the control - atypical time (815ms). Neither of these differences between responses to typical and atypical words in the control condition are statistically above chance. The difference between the experiment reported here and that of Whitney & Kellas can be interpreted as the result of a speed accuracy trade off, supported by the fact that corresponding error rates for control - typical conditions in Whitney & Kellas and this experiment are 7% and 1.1% (the highest and one of the lowest) respectively. As this difference occurred in the control condition, which is used as a plumb-line to judge the effects of context it may have biased the results of Whitney & Kellas in a positive direction and/or those of this experiment in a negative direction.

TARGET TYPE	SENTENCE PRIME		
	Control	Typical	Atypical
Typical	890 (7)	985 (5)	979 (3)
Atypical	910 (4)	930 (2)	930 (3)

TABLE 2
Mean Color-Naming Latencies and Percentage Error Rates (in Parenthesis) by Prime Type and Typicality.

In Whitney & Kellas, 1984.

More interestingly, in theoretical terms, are the very large variations that were noted in subjects responses to different words and sentences. Even the controls of asking subjects which category members were suggested by the sentences used was not effective in predicting response times. It is apparent that some more complex process is coming into play with the introduction of context than changes in the similarity structures of concepts. It may be that if context is not working in a bottom-up way attempts to understand it cannot succeed by building theories from the bottom up either.

In summary, this experiment did not confirm the conclusions of Whitney & Kellas (1984) suggesting that the effects they reported are not robust enough to generalize to different experimental tasks and context sentences. The adequacy of the amended Stroop task as an appropriate experimental paradigm in which to study online context effects must thus be reassessed. The large variation in subjects' response times in this experiment suggests that the Stroop paradigm is aimed at too simplistic a level to yield relevant information about context effects.

APPENDIX 3

Sentences with Typical/Modified Targets for Experiment 2

"You found it" said Andy, noticing the broken tool in the workshop.

HAMMER BROKEN HAMMER

Although it was borrowed, Jill thought that the old clothing would be adequate.

CLOTHES OLD CLOTHES

Although it was on loan, Kathy used the piece of antique furniture all the time.

CHAIR (swivel) ANTIQUE CHAIR

"They are all there" said Linda as her mother viewed the piled fruit in the hall.

ORANGE PILED ORANGES

Attempting to complete the job, Dick moved the loaded vehicle across the site.

LORRY LOADED LORRY

"It's over there" directed Mark, hoping Harry would find the hanging tool where he left it.

HAMMER HANGING HAMMER

Helen put it aside because she was in a hurry, even though the half-eaten fruit appealed to her.

APPLE HALF-EATEN APPLE

Though Sam had it with him he hoped that the harnessed weapon wouldn't be necessary.

GUN HARNESSED GUN

The woman screamed "Help!" as she stared at the bloody weapon in his hand.

KNIFE BLOODY KNIFE

The teenager had already acquired some small articles of modern furniture of her own.

CHAIR (dining-room type)   MODERN CHAIR

APPENDIX 4

Sentences with Typical/Atypical Referent Targets

for Experiment 3

He won the fairground fruit in a competition.

PEAR (T) COCONUT (AR)

Keith was not allowed to participate because his parents thought the cruel sport was unhealthy.

FOOTBALL (T) FOXHUNTING (AR)

The young rioter spotted a convenient weapon in the street and hurled it at the policeman.

PISTOL (T) BOTTLE (AR)

*The state authorities used their humane weapon to execute criminals.

RIFLE (T) ELECTRIC CHAIR (AR)

*The retriever brought back the scavenger bird that the hunter had seen.

ROBIN (T) VULTURE (AR)

*The heavy tool the fireman used to break into the locked room was a last resort.

SAW (T) AXE (AR)

*Although it was unusual the pale vegetable was served with the meal.

SPINACH (T) TURNIP (AR)

Mark got out his Scottish instrument to play at the lively parade.

GUITAR (T) BAGPIPES (AR)

Nick thought that the colourful bird he saw at the zoo was very tame.

DOVE (T) PARROT (AR)

Because the cheap vegetable was used the meal went further.

CARROT (T) POTATO (AR)

*Excluded in the reanalysis of Experiment 4.